

جامعة اليرموك

كلية التربية

قسم علم النفس الإرشادي والتربوي



"دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار"

(دراسة مقارنة)

Accuracy of Item Parameters and Ability Estimated Parametric and Non-Parametric Models of Item Response Theory
"Comparison Study"

إعداد

حسين عبد النبي القيسي

باشراف الأستاذ الدكتور

ساري سليم سواقد

حقل التخصص/ القياس والتقويم

دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار (دراسة مقارنة)

إعداد

حسين عبد النبي القيسي

ماجستير القياس والتقويم، جامعة مؤتة 2008م

قدمت هذه الأطروحة استكمالاً لمتطلبات الحصول على درجة دكتوراه الفلسفة تخصص القياس والتقويم في جامعة اليرموك، إربد، الأردن.

وافق عليها

ساري سليم سواقد.. مشرفاً ورئيساً

أستاذ في القياس والتقويم، جامعة مؤتة

غازي ضيف الله رواقة.. عضواً

أستاذ في مناهج التربية المهنية وأساليب تدريسها، جامعة اليرموك

أحمد يوسف قواسمة.. عضواً

أستاذ في القياس والتقويم، جامعة اليرموك

نضال كمال الشريفين.. عضواً

أستاذ مشارك في القياس والتقويم، جامعة اليرموك

يحيى حياتي نصار.. عضواً

أستاذ مشارك في القياس والتقويم، الجامعة الهاشمية

تاريخ مناقشة الأطروحة: 2013/8/4

الإهداء

إلى جامعتين، عينين دافقتين بالنور والعطر
إلى جناحين خافقين بكل المعاني التاريخية الأردنية الإنسانية السامية.. إلى جامعة مؤتة وجامعة اليرموك وإلى كل من ساهم في إثرائي وتسهيل نجاحي وبلوغي هذه المرحلة من المعرفة.

أقدم هذا العمل..

واعتز برفقة كان لها كل الأثر الطيب في وصولي هنا...

شكر وتقدير

الحمد لله والصلاة والسلام على سيدنا محمد وآله وسلم

إنه من تمام الأعمال وكمال الأخلاق، الاعتراف لذوي الفضل بفضلهم وشكرهم وتقديرهم، لذا فإنني أتقدم بالشكر الجزيل إلى أستاذي الفاضل الأستاذ الدكتور ساري سليم سواقد، الذي أشرف على هذه الرسالة فكان لي عوناً وسنداً من خلال توجيهاته الحكيمة التي أخرجت هذا العمل إلى النور فله كل الاحترام والتقدير.

كما أتقدم بالشكر الجزيل إلى أعضاء لجنة المناقشة الأفاضل: الأستاذ الدكتور أحمد يوسف قواسمة والأستاذ الدكتور غازي ضيف الله رواقة والدكتور يحيى حياتي نصار والدكتور نضال كمال الشريفين.

وأخيراً أسأل الله تعالى أن ينفعنا بما علمنا وأن يعلمنا ما ينفعنا والحمد لله رب العالمين.

المحتويات

العنوان | رقم الصفحة

فهرس الجداول

فهرس الأشكال

رقم الشكل | الموضوع | رقم الشكل

الملخص

القيسي، حسين عبدالنبي. دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية باختلاف حجم العينة وطول الاختبار. أطروحة دكتوراه، جامعة اليرموك،2013. (المشرف: أ. د. ساري سليم سواقد)

هدفت الدراسة إلى مقارنة دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية استجابة الفقرة البارامترية واللابارامترية، باختلاف حجم العينة، وطول الاختبار، اعتمادا على مؤشري دقة القياس التحيز (BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، ولتحقيق أغراض الدراسة، تم توليد قدرات لأفراد عينات حجمها 100، و250، و500، و1000، فرد من توزيع طبيعي بوسط حسابي يساوي صفراً، وانحراف معياري مقداره (1). وبالاعتماد على معلمة القدرة تم توليد اختبارات وفقاً للتصنيف الآتي (20، و40، و60) فقرة تحت افتراض التوزيع الطبيعي لصعوبة الفقرات بوسط حسابي مقداره (0)، وانحراف معياري مقداره (1)، والتوزيع المنتظم لمعلمة التمييز بواقع قيمة ابتدائية (0.4)، وقيمة نهائية (1.2)، والتوزيع المنتظم لمعلمة التخمين بواقع قيمة ابتدائية (0.2) وقيمة نهائية (0.3) على افتراض أن الاختبارات من نوع الاختيار من متعدد، وله أربعة بدائل،تلائم النموذج ثلاثي المعلم باستخدام برنامج توليد البيانات WinGen.

وللإجابة عن اسئلة الدراسة، استُخدمت برمجية BILOG-MG؛ لتقدير معالم الفقرة، والقدرة باستخدام طريقة الأرجحية العظمى الهامشية MML البارامترية، ومن ثم تمت مقارنة المعالم المقدرة مع المعالم الحقيقية (المولدة) باستخدام مؤشري التحيز (BIAS) والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، ثم تم حساب معامل ارتباط بيرسون بين المعالم الحقيقية والمقدرة كمؤشر ثقة، واستخدمت برمجية TESTGRAF؛ لتقدير معالم الفقرة والقدرة باستخدام طريقة تنعيم النواة KS اللابارامترية، ومن ثم تمت مقارنة المعالم المقدرة مع المعالم

الحقيقية (المولدة) باستخدام مؤشري التحيز (BIAS) والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، ثم تم حساب معامل ارتباط بيرسون بين المعالم الحقيقية والمقدرة كمؤشر ثقة.

وقد أظهرت نتائج الدراسة أن قيم معاملات الارتباط بين معلمة القدرة (θ) الحقيقية ومعلمة القدرة (θ) المقدرة بالطريقة البارامترية والمقدرة بالطريقة اللابارامترية، قد كانت موجبة ودالة إحصائيا، ويمكن تصنيفها بالعالية، كما أظهرت نتائج الدراسة أن قيم معاملات الارتباط بين المعالم الحقيقية (a, b, c) والمعالم المقدرة بالطريقة البارامترية قد كانت موجبة ودالة إحصائيا، ويمكن تصنيفها بالعالية، وهي أكبر من معاملات الارتباط بين المعالم الحقيقية والمعالم المقدرة بالطريقة اللابارامترية في معظم حالات الدراسة تقريبا، وفيما يتعلق بمؤشرات الدقة في القياس، فقد أشارت النتائج إلى أن قيم التحيز وقيم الجذر التربيعي لمتوسط مربعات الخطأ لتقديرات معالم الفقرة والقدرة بالطريقة البارامترية أقل منها في الطريقة اللابارامترية في معظم حالات الدراسة تقريبا. كما يلاحظ من خلال نتائج تحليل التباين الثلاثي للقياسات المتكررة لتقديرات معلمة الفقرات (a) ومعلمة القدرة (θ) وجود فروق دالة إحصائيا عند مستوى الدلالة α =0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) ومعلمة القدرة (θ) المقدرة، تعزى لتفاعل نوع النموذج المستخدم (بارامتري، و لابارامتري) مع متغيري (طول الاختبار، وحجم العينة). ويلاحظ من خلال نتائج تحليل التباين الثلاثي للقياسات المتكررة لتقديرات معلمتي الفقرات (b , c) عدم وجود فروق دالة إحصائيا عند مستوى الدلالة α =0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمتي الفقرات (b , c) المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة).

وأما فيما يتعلق بمؤشر الدقة في القياس RMSE، فقد أشارت النتائج إلى أن قيم متوسطاتها لطريقة التقدير البارامترية فيما يخص معالم الفقرات (a , b , c) ومعلمة القدرة (θ) كانت أقل من متوسطاتها لطريقة التقدير اللابارامترية في حالات الدراسة كلها، وأن هذا الفرق يقل كلما زاد حجم العينة وطول الاختبار.

أما بالنسبة إلى مدى التوافق كمؤشر ثقة، فقد أشارت النتائج إلى أن مدى التوافق بين قيم تقديرات معلمة القدرة (θ)، ومعالم الفقرات (a , b , c) لفقرات الاختبار ذي البيانات المولدة المقدرة حسب الطريقة البارامترية وبين قيم تقديرات معلمة القدرة الحقيقية وبين قيم تقديرات معالم الفقرات قد كانت أكبر من مدى التوافق بين قيم تقديرات معلمة القدرة (θ)، وقيم تقديرات معالم الفقرات (a , b , c) لفقرات الاختبار ذي البيانات المولدة المقدرة حسب الطريقة اللابارامترية.

الكلمات المفتاحية (دقة التقدير، ومعالم الفقرة والقدرة، ونماذج نظرية استجابة الفقرة البارامترية واللابارامترية).

الفصل الأول

خلفية الدراسة وأهميتها

الفصل الأول

خلفية الدراسة وأهميتها

مقدمة

يعد القياس من القضايا الرئيسية التي تهتم بها العلوم الطبيعية، والعلوم السلوكية الإنسانية على حد سواء؛ وذلك لأن القياس Measurement يستخدم الوصف الكمي للتعبير الموضوعي عن الظواهر المختلفة، وكلما كان القياس موضوعيا، زادت الدقة في فهم الظاهرة موضوع القياس والتنبؤ بها، من خلال بناء معادلات رياضية (قوانين) تحكم العلاقة الكمية بين المتغيرات المختلفة للظاهرة، الأمر الذي يؤدي إلى الاستخدام المقنن للظواهر والضبط والتحكم بها وهو الغاية النهائية للعلم، ومن دون ذلك يصبح العلم مقتصرا على الملاحظات الانطباعية العابرة، والتصنيف الكيفي، والأحكام الذاتية.

ويقصد بالموضوعية في القياس النفسي والتربوي عدم تأثر قياس الظاهرة السلوكية باختلاف الأداة المستخدمة في القياس، وألا يتأثر باختلاف الأفراد الذين يتم تقدير السمة لديهم من خلال هذه الأداة، وأن يتم تدريج الأداة بوحدة قياس تتوافق مع تدريج مستويات الظاهرة السلوكية موضوع القياس، كما يحدث في قياس الظواهر الفيزيائية .

وعلى الرغم من أن مشكلات القياس لا تختلف كثيرا باختلاف العلوم، إلا أن الباحثين في العلوم السلوكية أكثر اهتماماً بالأسس المنطقية للقياس من نظرائهم في العلوم الطبيعية، كون السمات النفسية والتربوية سمات كامنة يستدل عليها من خلال السلوك المرتبط بها الأمر الذي يجعل قياسها غير مباشر؛ أي من خلال السلوك المرتبط بهذه السمة وغير المرتبط بسمات أخرى، وتتميز الظواهر السلوكية بالتعقيد، وتعدد المتغيرات وتشابكها، فقياس السمات

والخصائص الإنسانية، مثل: التحصيل، والذكاء، والاستعداد والميول، والاتجاهات، والشخصية، يختلف عن قياس الخصائص الفيزيائية (علام، 2005).

ويتصف قياس السمات النفسية والتربوية (السلوكية) بالخصائص الآتية:

- غير مباشر Indirect؛ أي أنه لا يتم قياس السمة، أو مقدار ما يمتلكه الفرد من تلك السمة بشكل مباشر، بل من خلال الأداء على مواقف لها علاقة بالسمة ذاتها، أي أنه يمكن قياس السمات النفسية والتربوية من خلال السلوك الظاهري القابل للملاحظة والقياس والمرتبط بالسمة موضوع القياس.
- غير تام Incomplete؛ أي لا يتم قياس جميع السلوكيات المكونة للمجال السلوكي للسمة، بل يتم قياس عينة من هذا المجال .
- نسبي Relative، أي لا يمكن تفسير درجة الفرد على الاختبار النفسي أو التربوي دون مقارنتها بمستوى أداء المجموعة التي ينتمي إليها ذلك الفرد، أو بمقارنة مستوى اداء الفرد بمعيار معين أو بمحك (Criterion) او بمعيار (Norm) (النبهان، 2004) .

وهذه الخصائص تجعل احتمال الوقوع في خطأ القياس للسمات النفسية والتربوية أكثر منه في حالة السمات الفيزيائية (القياس المادي)، الأمر الذي أثار اهتمام المختصين في القياس النفسي والتربوي، وأدى ذلك إلى الحاجة إلى نظريات تمكن افتراضاتها من إيجاد طرق لتقدير درجة دقة القياس. ونتج عن هذه النظريات نماذج إحصائية سيكومترية، لتقدير دقة القياس للاختبارات النفسية والتربوية، وتفسير نتائجها، وتنقسم النماذج الإحصائية السيكومترية إلى قسمين رئيسين، هما:

- **نظرية الاختبار أو ما يسمى بالنظرية الكلاسيكية CTT (Classical Test Theory):**

التي ساد استخدامها لفترة غير قصيرة على بناء أنواع الاختبارات النفسية والتربوية وتطويرها، ويتم التحقق من دقة القياس في نظرية الاختبار عن طريق مفهوم الثبات (ثبات نتائج أداة القياس)، إذ يعرف الثبات في النظرية الكلاسيكية في القياس بأنه "مدى التغير الناتج عن الأخطاء العشوائية في درجات شخص ما عند تطبيق عدة اختبارات متوازية عليه أو عند تطبيق الأداة نفسها مرات عديدة" تحت نفس لظروف (Ghisiell, 1981). إلا أن هذه النظرية ظهرت فيها جوانب قصور (Shortcomings) تمثلت بنواحي عدة مما أدى إلى البحث عن منحى جديد ومغاير للنظرية الكلاسيكية لمعالجة المشاكل التي تعاني منها، والتي أشار إليها هامبلتون وسوامنثان (Hambleton & Swaminathan; 1985) على النحو الآتي:

- عدم الموضوعية في القياس، التي تظهر من خلال عدم استقلالية قيم إحصاءات فقرات الاختبار (Item Statistics) (الصعوبة والتمييز) عن خصائص عينة المفحوصين، وإحصاءات المفحوصين (Person Statistics) غير مستقلة عن عينة فقرات الاختبار.
- يتأثر موقع المفحوص على متصل القدرة أو السمة باستخدام فقرات اختبار معين بالأوضاع والظروف التي تم فيها تطبيق الاختبار نفسه أو اختبار موازي له.
- من المفاهيم الأساسية للنظرية الكلاسيكية مفهوم الثبات، الذي يعرف بدلالة الاختبارات المتوازية أو إعادة الاختبار، ولكن من الصعب تحقيق مفهوم "المتوازية"، أو الشروط الواجب توافرها عند إعادة الاختبار من الناحية العملية.

- لا تزودنا هذه النظرية بأي أساس للتنبؤ في كيفية أداء المفحوص على فقرة اختبار ما؛ بمعنى أنها لا تستطيع الإجابة عن السؤال، ما احتمال أن يجيب مفحوص ما إجابة صحيحة عن فقرة اختبار محددة.
- تفترض هذه النظرية أن تباين أخطاء القياس هو نفسه لمستويات القدرة جميعها.

ولتلافي جوانب القصور السابقة، وحل مشكلة الموضوعية، والدقة في القياس، ظهر الاتجاه الجديد في القياس، الذي يقوم على نماذج رياضية تعتمد على نظرية الاحتمالات، وهي نماذج عرفت بنماذج السمات الكامنة (Latent Trait Models)، واستخدام هذه النماذج في تحليل الفقرات يُمَكن من الحصول على:

- إحصاءات للفقرة لا تعتمد على خصائص المفحوصين.
- درجات تعبر عن قدرة المفحوصين (أو مقدار ما يمتلك الفرد من السمة موضوع القياس) لا تعتمد على الخصائص السيكومترية (الصعوبة والتمييز والتخمين) لفقرات أداة القياس (الاختبار أو المقياس).
- نموذج لنظرية يوفر الأساس لربط احتمال الحصول على الإجابة الصحيحة لفقرات الاختبار بمستويات القدرة للأفراد.
- نماذج لا تقوم على افتراضات غير ممكنة (صعب الدفاع عنها).
- نماذج من الاختبار لا تتطلب التوازي الصارم لتقدير الثبات (الدقة) الذي يصعب تحقيقه عملياً. بالإضافة إلى أن هذه النماذج (المعاصرة) تقدر خطأ قياس عند كل مستوى من مستويات القدرة بعكس النماذج الكلاسيكية التي تفترض خطأ قياس واحد (علام، 2005).

وتنقسم نماذج نظرية استجابة الفقرة إلى نوعين رئيسيين، الأول: يعرف بنماذج نظرية استجابة الفقرة البارامترية (Parametric Item Response Theory Models) (PIRT)،

حيث يكون شكل دالـــة اســـتجابة الفقـــرة (IRF) (Item Response Function) محـــددة (لوجستية الشكل)، بينما يعرف النوع الثاني بنماذج نظرية اســتجابة الفقــرة اللابارامتريـــة (Non-Parametric Item Response Theory Models) (NIRT)، التي لا تحدد شكل محدد لدالة استجابة الفقرة (فقط أن تكون غير متناقصة)، ولا تفترض أي شكل سابق، وكمـــا يشير فان ديـــر لنـــدن وهـــامبلتون (Van der Linden & Hambleton,1997) يمكـــن الافتراض أن هذه الدوال أقرب لدوال الاستجابات الحقيقية من تلك التـــي تعطيهـــا النمـــاذج البارامترية؛ لأنها تعتمد على افتراضات أقل حول النموذج الحسابي.

وبالرغم من الاختلافات بين هذه النماذج، إلا أن النماذج البارامتريـــة واللابارامتريـــة تشترك في كثير من الاستخدامات التطبيقية كتطوير المقاييس المختلفة، وتكون المفاضلة بينهما في ضوء طبيعة البيانات المستخدمة، وأيهما يعطي تقديراً أدق، وعلى ماذا يعتمد ذلك.

دقة القياس في نماذج استجابة الفقرة البارامترية واللابارامترية:

تعدُّ نماذج استجابة الفقرة البارامترية حالـــة خاصـــة مـــن نمـــاذج اســـتجابة الفقـــرة اللابارامترية، إذ يمكن تحليل بيانات النماذج البارامترية وفق المعالجات الإحـــصائية لتقـــدير معالم الفقرة والقدرة باستخدام النماذج اللابارامترية، ولا يجوز العكـــس، وذلـــك أن النمـــاذج اللابارامترية تستطيع تحليل بيانات على مستوى القياس الرتبي بعكس النمـــاذج البارامتريـــة، التي تشترط وقوع البيانات على مستوى القياس الفئوي على الأقل؛ أي أنه إذا كان من الممكن قياس السمة على مستوى أعلى، فإننا نستطيع قياسها على مستوى أدنى والعكـــس لا يجـــوز، وذلك بسبب خاصية الهرمية في مستويات القياس، ولكن قـــد يكـــون هنـــاك اخـــتلاف بـــين النموذجين في دقة التقديرات التي يعطيها كلاهما (Sijtsma & Hemker, 2000).

وأن تحديد دقة التقدير تتم من خلال ما يعرف بـ" الخطأ المعياري للتقدير"، وهو معكوس الدقة بمعنى كلما زادت المعلومات التي يقدمها الاختبار عن قدرات المفحوصين، زادت الدقة. وبما أن دقة التقدير تتأثر بحجم عينة المفحوصين وعدد الفقرات، فإن اختيار حجم العينة وطول الاختبار المناسبين للتقدير أمر مهم جدا، وقد بين هلن وليساك ودراسجو (Hulin, lissak, and Drasgow, 1982) المشار إليهم في (Crocker and Algena, 1986) أن الدقة في تقدير معالم الفقرة الثلاثة في نماذج السمات الكامنة، تزداد بزيادة حجم العينة من (200) إلى (500) ثم إلى (1000) على الترتيب، إلا أنه لم يتبين أي تحسن في دقة التقدير عند زيادتها إلى (2000).

وتختلف النماذج البارامترية واللابارامترية عن بعضها بعضاً في أسلوب التقدير، وحجم العينة، وطول الاختبار المناسبين للتقدير، حيث إن النماذج اللابارامترية تحتاج إلى حجم عينة وطول اختبار أقل من نظيرتها البارامترية، وأدعى رامسي (Ramsay, 1991) أن عددا قليلا من المفحوصين لا يزيد عن 100 وعدد فقرات الاختبار 20 هما المطلوبان لتقدير (ICCs). إضافة إلى أن السمات النفسية والتربوية لا تقع على مستوى القياس الفئوي، بل تقع على مستوى قياس أرقى من الرتبي وأقل من الفئوي (بين الرتبي والفئوي). من هنا جاءت فكرة هذه الدراسة، وهي مقارنة دقة تقدير معالم الفقرات والقدرة لنماذج نظرية استجابة الفقرة البارامترية واللابارامترية، باختلاف حجم العينة، وطول الاختبار، محاولاً الإجابة عن السؤال العام وهو "ما مدى دقة تقدير النموذجين لمعالم الفقرة والقدرة"؟

وتكتسب الطرق الإحصائية اللابارامترية جاذبية خاصة لدى العاملين في مجال البحث العلمي، ومن أسباب ذلك أنها لا تتطلب افتراضات قوية حول توزيع المجتمع كما هو الحال عادة في طرق الإحصاء البارامتري، إضافة إلى أن بعضها يمكن استخدامه حتى في حالة

البيانات المقاسة بمقاييس دنيا، كالمقياس الاسمي والرتبي. ويشير (عوده، 2000) نقلاً عن بوفام (Popham and Sirotnik, 1973) أن الاختبارات اللابارامترية بصورة عامة أكثر قوة من الاختبارات البارامترية، إذ تميل الإحصائيات البارامترية أكثر من اللابارامترية لرفض الفرضية الصفرية، كما أن الاختبارات اللابارامترية أسهل في طريقة تطبيقها.

الطرق البارامترية واللابارامترية

يطلق مصطلح الطرق البارامترية Parametric Methods على الطرق أو المعالجات الإحصائية التي تستند على توزيعات ذات معالم (وسط حسابي وانحراف معياري) محددة للبيانات المراد تحليلها؛ أي التي تستخدم للاستدلال على بارامتر وتستند إلى افتراضات قوية حول توزيع المجتمع مثل التوزيع الطبيعي وتجانس التباين. بالإضافة إلى ذلك، فإن المتغيرات التابعة يفترض أن تكون مقاسة بمقياس فئوي على الأقل، بينما يستخدم مصطلح الطرق اللابارامترية Nonparametric Methods أو طرق غير معتمدة على التوزيع Distribution—free methods على الطرق التي لا تستخدم للاستدلال على بارامتر أو لا تستند إلى افتراضات قوية حول توزيع المجتمع.

مزايا وعيوب الطرق اللابارامترية

تتميز الطرق اللابارامترية كما يشير (الدردير، 2006) بمزايا عدة مقارنة بالطرق البارامترية، وهي:

1- بساطة الافتراضات وسهولة تحقيقها

لعل أهم مزايا الطرق اللابارامترية أنها لا تتطلب افتراضات صعبة أو قوية حول البيانات، كما هو الحال عادة في الطرق البارامترية. وهذه ميزة كبيرة لأن الكثير من مستخدمي الطرق الإحصائية قد لا يعرفون ما إذا كانت الافتراضات التي تقوم عليها متحققة

في بياناته، أو قد يعرف أن بعضاً منها أو كلها غير متحققة. واستخدام طريقة بارامترية دون التأكد من تحقق الافتراضات التي بنيت عليها يؤدي لنتائج غير دقيقة. أما الطرق اللابارامترية، فإن الافتراضات القليلة التي قد تتطلبها (بعض الطرق اللابارامترية لا تتطلب أي افتراضات حول التوزيع) تكون عادة عامة، ومن النوع الذي يتحقق في معظم التطبيقات، مثل افتراض أن التوزيع متصل، فمثلاً اختبار مان وتني للمقارنة بين عينتين مستقلتين عندما تكون البيانات عددية بطبيعتها، وهذا الاختبار هو البديل اللابارامتري لاختبار (t) البارامتري للبيانات المستقلة للفرق بين متوسطين من عينتين مستقلتين – وهو اختبار بارامتري شائع الاستخدام يفترض لاستخدامه أن يكون توزيع المجتمعين طبيعي وأحياناً بتباينين متساويين.

2- إمكانية التطبيق على البيانات الوصفية والترتيبية

في كثير من الدراسات، وخاصة في مجال العلوم الاجتماعية تكون البيانات ذات طبيعة وصفية أو ترتيبية، فمن البيانات الوصفية تلك التي تمثل تصنيف الوحدات حسب معيار معين، مثلاً تصنيف أشخاص حسب جنسهم أو جنسيتهم. ومن البيانات الترتيبية– على سبيل المثال – ترتيب مجموعة من العمال حسب درجة حماسهم أو رضاهم عن عملهم. ولقد استحدثت معظم الطرق اللابارامترية لمعالجة مثل هذا النوع من البيانات. ذلك أن تطبيق الطرق البارامترية على بيانات وصفية أو ترتيبية يؤدي عادة لنتائج يصعب أو يستحيل تفسيرها أو إعطاؤها معنى.

3- سهولة جمع البيانات وتحليلها

بما أن البيانات المستخدمة في الطرق اللابارامترية تكون عادة بمقاييس تصنيفية أو ترتيبية، مثل المقياس الاسمي أو الترتيبي، وبما أن معظم الطرق اللابارامترية لا تتطلب حجم عينة كبير أو حسابات معقدة، فإنه يمكن جمع البيانات وتحليلها بسرعة أكبر. فمثلاً إذا كانت

البيانات ستقاس بالمقياس الاسمي وحجم العينة صغير جداً، فإن جمع البيانات لن يستغرق سوى القليل من الوقت.

ومن المزايا الأخرى للطرق اللابارامترية أنها في معظم الحالات لا تتأثر بعدم تحقق الافتراضات التي تقوم عليها، التي هي في الأصل افتراضات ضعيفة وقليلة كما ذُكر. وهذه الخاصية تمثل ما يطلق عليه في الإحصاء الاستدلالي المنعة (Robust)، وتشير إلى قدرة الأداة الإحصائية للصمود في وجه عدم تحقق افتراضات تقوم عليها. كذلك فإن بعض الاختبارات اللابارامترية أقوى More Powerful من شبيهاتها البارامترية خاصة في حالة العينات الصغيرة.

نظرية استجابة الفقرة (Item Response Theory)

تعد نظرية استجابة الفقرة (IRT) في القياس ثورة في القياس النفسي والتربوي (علام، 2000). إذ تفترض أنه يمكن التنبؤ بأداء الأفراد، أو يمكن تفسير أدائهم في اختبار نفسي أو عقلي، في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات، وتحاول تقدير مقادير السمات عند الأفراد، واستخدام هذه المقادير في التنبؤ بأداء الأفراد على الاختبار والفقرة، ونظرا إلى صعوبة ملاحظة هذه السمات بصورة مباشرة، فإنه يجب تقديرها أو الاستدلال عليها من أداء الأفراد في مجموعة من فقرات الاختبار، ولهذا السبب تسمى بالسمات الكامنة (Hambleton & Swaminathan, 1985).

ويشير كليف وكيتس (Cliff & Keats, 2003) إلى أن نماذج نظرية استجابة الفقرة تختلف باختلاف البيانات، وتنقسم بشكل رئيسي إلى نوعين أساسيين، هما النماذج البارامترية والنماذج اللابارامترية لنظرية استجابة الفقرة، إلا أنها جميعها تقوم على فكرة وجود تدريج كامن (latent scale) لقدرة معينة أو سمة شخصية، وأن احتمال الإجابة الصحيحة أو

(الموافقة) على الفقرة تتغير بطريقة بسيطة عبر ذلك التدريج الكامن، حيث تزداد هذه الاحتمالية بازدياد قدرة المفحوص. وقد تم عرض دوال رياضية مختلفة حول ماهية شكل دالة استجابة الفقرة التي تعبر عن العلاقة بين قدرة المفحوص واحتمالية الإجابة الصحيحة، وتم اقتراح الشكل اللوجستي، الذي يمثل المنحنى البارامتري منذ سنوات عدة كخيار جيد.

بينما تأخر الحديث عن المنحى اللابارامتري حتى بداية السبعينات من القرن العشرين. وقدمت نظرية استجابة الفقرة معايير مهمة تفيد في عملية اختيار وتقدير معالم الفقرات، الأمر الذي ميزها عن غيرها من النظريات، فهي تعتمد على دالة استجابة الفقرة بشكل أساسي في هذه العملية بالإضافة إلى الاستفادة من دالة معلومات الفقرة (Item Information Function)، ودالة معلومات الاختبار (Test Information Function)، الأمر الذي طور عملية بناء الاختبارات والتنبؤ بنتائجها (Lord, 1977).

ويشير هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) إلى أن نظرية استجابة الفقرة (IRT) تتمتع بمجموعة من الخصائص التي ميزتها عن غيرها من النظريات، وهذه الخصائص هي:

1- استقلالية خصائص معالم الفقرة عن عينة المفحوصين، التي استخدمت لتقدير هذه البارامترات (Person-Free).

2- استقلالية تقدير قدرات المفحوصين عن خصائص عينة الفقرات التي تم تطبيقها (item-free).

3- توفير مؤشرات إحصائية لتقدير دقة قياس قدرة كل مفحوص.

نماذج نظرية استجابة الفقرة البارامترية

نتجت عن هذه النظرية مجموعة من النماذج تعرف باسم نماذج السمات الكامنة، وفيما يلي توضيح لأشهر النماذج المستخدمة باختلاف عدد البارامترات، إذ تختلف هذه النماذج في الصورة الرياضية التي تمثل المنحنى المميز للفقرة (Hambleton & Swaminathan, 1985):

أولا. النموذج اللوجستي أحادي البارامتر (نموذج راش) One Parameter Logistic Model (Rash Model):

أول من نشر هذا النموذج هو عالم الرياضيات الدنماركي جورج راش (Georg Rash)، ويفترض النموذج أن الفقرات جميعها تميز بنفس القدر بين المفحوصين، لكنها تتباين فقط في صعوبتها، ويتميز النموذج بسهولة التعامل معه مقارنة بالنماذج الأخرى، إذ يعد أقل نماذج السمات الكامنة في متطلباته، ويتخذ المنحنى المميز للفقرة في هذا النموذج المعادلة الرياضية الآتية:

$$P^{i(\theta)} = \frac{1}{1 + e^{-Da(\theta - b)}}$$

$P_i(\theta)$:احتمال الإجابة الصحيحة عن الفقرة i من مفحوص قدرته (θ)، D : ثابت ويدل على معامل التدريج، bi بارامتر صعوبة الفقرة i، θ : قدرة الفرد

والمتغير التابع في نموذج راش البسيط هو احتمال أن فرد (j) يجيب إجابة صحيحة على فقرة اختبارية (i). أما المتغيرات الكامنة المستقلة، فهي درجة قدرة الفرد (θ_j)، وصعوبة الفقرة (b_i). وتضم المتغيرات المستقلة بعملية جمع، كما أن صعوبة الفقرة (b_i) تطرح من قدرة الفرد (θ_j). وعلاقة هذا الفرق بالاستجابة للفقرة يعتمد على اختيار المتغيرات المستقلة

التي يتم نمذجتها في صيغة رياضية احتمالية، وهي دالة ترجيح لوغاريتمي غير خطية Nonlinear Logistic Function (علام، 2005).

ويتميز نموذج راش بسهولة التعامل معه؛ لأنه يستخدم عدداً قليلاً من البارامترات مقارنة بالنماذج الأخرى، ويمكن الحصول على قدرة الفرد بمعرفة صعوبة الفقرة، واحتمال الإجابة الصحيحة عليها، وكذلك الحصول على صعوبة الفقرة من خلال معرفة قدرة الفرد، واحتمال الإجابة الصحيحة على الفقرة. بالإضافة إلى أن كلاً من صعوبة الفقرة ومستوى القدرة مستقلان، وهذا يسمح بتقدير مستقل لكل منهما؛ أي أن تقدير القدرة مستقل عن الفقرة المستخدمة في التقدير، وكذلك يمكن تقدير صعوبة الفقرة بغض النظر عن قدرة أفراد العينة المستخدمة، وهذا الاستقلال في التقدير لهذه البارامترات هو جوهر وأهم خاصية تقريباً في نظرية القياس، وقد أسماها راش بالموضوعية المحددة Specific Objectivity.

ثانيا. النموذج اللوجستي ثنائي البارامتر Two Parameter Logistic Model

يقوم هذا النموذج على افتراض أن الفقرات تختلف في صعوبتها وتمييزها، وغياب عامل التخمين، ويعد النموذج أكثر واقعية من النموذج اللوجستي أحادي البارامتر؛ لأنه من الصعب إيجاد مجموعة من الفقرات لها نفس القدرة التمييزية على مستويات مختلفة من القدرة، لذلك تضمنت الصيغة الرياضية لهذا النموذج بارامتر تمييز الفقرة، وأصبح يشتمل النموذج على بارومترين (الصعوبة والتمييز) والمعادلة الرياضية لهذا النموذج هي:

$$Pi(\theta)=\frac{1}{1+e^{-Dai(\theta-b)}}$$

ويفترض هذا النموذج أن الفقرات تختلف عن بعضها بعضاً في التمييز والـصعوبة، كما يفترض أن الأفراد ذوي القدرة المنخفضة، لا يحصلون علـى إجابـة صـحيحة للفقـرة الصعبة؛ أي أن التخمين معدوم (Hambleton, 1989).

ثالثاً. النموذج اللوجستي ثلاثي البارامتر Three Parameter Logistic Model

إن نموذج راش، والنموذج ثنائي البارامتر لا يأخذان بعين الاعتبار أن احتمال الإجابة الصحيحة على مفردة اختبار من نوع اختيار من متعدد يكون أكبر بالنسبة للأفراد من ذوي القدرة المنخفضة؛ لأنهم ربما يلجؤون إلى التخمين العشوائي (علام، 2005).

ويعد النموذج ثلاثي البارامتر (3PL) امتدادا للنموذج اللوجستي ثنائي البارامتر، إذ يضيف بارامتراً جديدا للفقرة، وهو بارامتر التخمين (c)، ويشير إلى احتمال إجابة الفقرة إجابة صحيحة من المفحوصين ذوي القدرة المتدنية، ويسمى Pseudo Chance Level، والمعادلة الرياضية لهذا النموذج هي:

$$pi(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b\ )}}{1 + e^{-Da_i(\theta - b_i)}}$$

$P_i(\theta)$:احتمال الإجابة الصحيحة عن الفقرة i من مفحوص قدرته (θ)، D : ثابت ويدل على معامل التدريج، bi بارامتر صعوبة الفقرة i، ai : بارامتر تمييز الفقرة، ci : معلم تخمين الفقرة، θ : قدرة الفرد.

وهناك مجموعة من النماذج طورت لمعالجة بيانات متعـددة الأبعـاد، يـشار إليهـا بالنماذج المتعددة الأبعاد التعويضية، وغير التعويضية، إذ يفترض النمـوذج التعويـضي أن قدرات المفحوص تتفاعل لتنتج استجابة على فقرة ما، بحيث يأخذ هذا التفاعـل مـن قـدرة المفحوص على سمة كامنة ما، لتعويض النقص في قدرته على سمة كامنة أخرى، أما النماذج

غير التعويضية فتفترض أن قدرة المفحوص في سمة ما لا تعوض قدرته في سمة أخرى لتحصل الاستجابة (De Ayala , 1993).

افتراضات نماذج استجابة الفقرة البارامترية

ترتكز نماذج النظرية الحديثة في القياس على مجموعة من الافتراضات، الواجب توافرها في البيانات المستمدة من الاختبار، إذ يعتمد الاختيار المناسب للنموذج على تحقق هذه الافتراضات في البيانات، وتعد الافتراضات القائمة عليها نماذج نظرية استجابة الفقرة قوية وصارمة، وتقوم هذه النماذج بشكل عام على الافتراضات الآتية: (Hambleton & Swaminathan, 1985)

أولا: افتراض أحادية البعد Unidimensionality، ويعني وجود قدرة واحدة تفسر أداء الفرد في الاختبار.

تفترض نظرية السمات الكامنة وجود سمات كامنة عدة تفسر أداء المفحوصين على مجموعة من فقرات الاختبار، وهذه المجموعة من السمات الكامنة تعرف بأبعاد فضاء السمات الكامنة Dimensional Latent Space، وتعتمد أبعاد هذا الفضاء على عدد السمات التي ينطوي عليها أداء الأفراد في فقرات الاختبار، ويمكن أن تتعدد هذه الأبعاد، كما يمكن افتراض أن الفضاء الكامن أحادي البعد؛ أي افتراض أن فقرات الاختبار متجانسة، وتقيس قدرة أو سمة كامنة واحدة، فهناك خط انحدار واحد لجميع المفحوصين الذين استجابوا عن فقرات الاختبار.

ثانيا: افتراض الاستقلال الموضعي Local Independence، ومن معانيه أن استجابات المفحوص للفقرات المختلفة في الاختبار مستقلة عن بعضها إحصائيا عند مستوى قدرة معين؛ أي أن أداء المفحوص على فقرة ما لا يتأثر سلبا ولا إيجابا بأدائه على الفقرات

الأخرى، ويذكر هامبلتون وسوميناثان (Hambleton & Swaminathan, 1991) إلى أن افتراض الاستقلال الموضعي يكافئ افتراض أحادية البعد، ويعني ذلك أنه إذا تحقق افتراض أحادية البعد في المقياس، فإن المقياس يحقق افتراض الاستقلال الموضعي.

ثالثا: منحنى خصائص الفقرة (Item Characteristic Curve (ICC، وهو اقتران رياضي يربط بين احتمال استجابة الفرد استجابة صحيحة على فقرة الاختبار، وبين القدرة التي تقيسها مجموعة الفقرات التي تحتوي على تلك الفقرة، ويشير بوضوح إلى أن احتمال إجابة الفقرة إجابة صحيحة يزداد بازدياد قدرة المفحوص، بسبب أن المنحنى تراكمي صاعد، وتوصف هذه المنحنيات في نماذج الاختبارات المصممة لقياس سمة واحدة (أحادية البعد) بدلالة بارامتر واحد، أو بارامترين، أو ثلاثة بارامترات.

رابعا: انتفاء عامل السرعة Non-Speediness، ويعني أن عامل السرعة في الأداء، ليس له تأثير على احتمالية الاستجابة بشكل صحيح عن فقرات الاختبار؛ أي أن إخفاق الفرد في الإجابة عن فقرات الاختبار يعود إلى انخفاض قدرته، وليس إلى تأثير السرعة في الإجابة.

ويوجد مجموعة من الافتراضات خاصة بكل نموذج من نماذج استجابة الفقرة البارامترية، حيث إن النموذج الثلاثي يعتبر الحالة العامة للنماذج، وذلك لأنه يتم فيه تقدير بارامتر الصعوبة (b)، ومعلمة التمييز (a)، ومعلمة التخمين (c). أما النموذج الأحادي، فيتم فيه تقدير بارامتر الصعوبة (b) فقط، ويفترض أن التمييز متساوٍ لجميع الفقرات، ويساوي واحد، والتخمين (c) متساوٍ لجميع الفقرات، ويساوي صفر. أما النموذج الثنائي، فيتم فيه تقدير بارامتر الصعوبة (b)، ومعلمة التمييز (a)، ويفترض أن بارامتر التخمين (c) متساوٍ لجميع الفقرات، ويساوي صفر، وهذه الافتراضات يصعب تحقيقها (وجود مجموعة من الفقرات

تمييزها متساوٍ، ويساوي واحداً وأن احتمالية التخمين تساوي صفر للفقرات ذات نوع الإجابة الاختيار من متعدد).

ومن الممكن أن تظهر مشاكل النماذج اللوغاريتمية أحادية المعلم وثنائية المعلم عند تطبيقها على فقرات الاختيار من متعدد، أو الصح والخطأ، وذلك لأن هذه الصيغ جميعا تسمح بالإجابات الناتجة عن التخمين. والمشكلة في النماذج أحادية وثنائية المعلم أن قيمة Pi(θ) تقترب من الصفر، كلما انخفضت قيمة (θ). ومع ذلك يمكن توقع نسبة الإجابة للمفحوص ذي القدرة المنخفضة أكبر من الصفر؛ وذلك لأن بإمكان هذا المفحوص أن يخمن الإجابة الصحيحة، لذلك فإنه يمكن استخدام النموذج ثلاثي المعلم بسبب هذا العامل (Sijtsma, 2002).

ويعد النموذج اللوجستي ثلاثي البارامتر (Three Parameter Logistic Model (3PLm) المستخدم في هذه الدراسة أقل النماذج البارامترية تشددا، ذلك أنه يسمح أن تختلف فقرات الاختبار في صعوبتها وتمييزها. إذ من الصعب إيجاد مجموعة من الفقرات التي تميز بدرجة واحدة بين مستويات السمة، أو القدرة التي يقيسها اختبار معين، كما في النموذج أحادي البارامتر (One Parameter Logistic Mode) (نموذج راش). كذلك يفترض تأثر الإجابات بعامل التخمين، وهو المعلم الثالث الذي يتميز به عن النموذج ثنائي البارامتر (Two-Parameter Logistic Model)، الذي أضافه لورد، وأطلق عليه بارامتر الخط التقاربي الأدنى (Lower Asymptote Line)، أو بارامتر التخمين (Guessing Parameter)، وهذا المعلم يحدد احتمال أن يجيب فرد يفترض أن مستوى قدرته متدنٍ جدا، ومع هذا يجيب إجابة صحيحة عن فقرات الاختبار عن طريق التخمين، والحقيقة أن إضافة هذا المعلم الثالث يكون بمثابة تعويض عدم مطابقة المنحنيات المميزة للفقرات عند النهاية

السفلى لمتصل القدرة، نتيجة تأثير إجابات هؤلاء الأفراد بعامـل التخمـين (Lord, 1980; Hambleton ,1989).

ويعتمد احتمال إجابة الفرد عن الفقرة إجابة صحيحة في نماذج نظرية الاستجابة للفقرة، على قدرة المفحوص (θ)، والبارامترات الخاصة بالفقرة متمثلة بمعلم الصعوبة (b)، ومعلـم التخمين (c)، ومعلم التمييز(a)، ولكن هذه القيم احتمالية، وهذه البـارامترات تكـون غيـر معلومة، بينما إجابات الأفراد عن الفقرة تكون معلومة، لذلك فإن أساليب تقدير البـارامترات تهدف لتحديد قيمة (θ) لكل فرد، وكذلك قيم معالم الفقرة من إجابات الأفراد عليها (الشريفين، 2012).

نماذج نظرية استجابة الفقرة اللابارامترية

تعد نماذج نظرية استجابة الفقرة اللابارامترية نماذج إحصائية يمكن اسـتخدامها فـي القياس النفسي والتربوي لدراسة جودة المقياس وكفاءته من خلال تحليلات جتمـان. وتـستند على مفهوم التراكمية (Guttman Scalogram Analysis)، التي تفترض أن كل فقرة وكل مفحوص يمتلكان موقعا على متصل القدرة، حيث يجيب المفحوص على الفقرة إجابة صحيحة إذا كانت قدرته أعلى من صعوبة تلك الفقرة، مما يتيح التنبؤ بنمط استجابة المفحـوص مـن معرفة درجته الكلية على الاختبار، وقد لا يتحقق ذلك تجريبيا مما أدى إلى ظهور الفكرة التي جاءت بها النماذج الحديثة لنظرية استجابة الفقرة (البارامترية واللابارامترية) والقائمة علـى الاحتمالية لا التحديد، التي نصت على أن احتمالية الإجابة الصحيحة مرتفعة (لكـن ≠ 1) كلما ازدادت قدرة المفحوص، كما أنها قليلة (لكن ≠ 0) عند انخفـاض قـدرة المفحـوص (Sijtsma & Hemker, 2000).

وتنبع أهمية استخدام نماذج استجابة الفقرة اللابارامترية من أنها نماذج تسمح بتقدير معالم مهمة، مثل نسبة الإجابات الصحيحة، ومؤشرات الصعوبة، ومعاملات التدريج المختلفة، التي تشير إلى القوة التمييزية لفقرات الاختبار والاختبار ككل. وتضمن هذه النماذج ترتيب المفحوصين باستخدام درجاتهم على الاختبار بالرغم من وجود الخطأ العشوائي، إذ أنها تعتبر أن هذه الدرجات تعكس رتب القدرة (θ).

ويشير سيجتسما وموليناز (Sijtsma & Molenaar, 2000) أن النماذج اللابارامترية تقوم على مجموعة من الافتراضات، التي تعد أقل تشدداً من تلك التي تقوم عليها النماذج البارامترية وهي كالآتي:

1- أحادية البعد (Unidimensionality): الاستجابات على الفقرات تتبع متغير كامن أحادي البعد يرمز له بالرمز (θ).

2- الاستقلال الموضعي (Local Independence): احتمالية الإجابة على أي فقرة غير مرتبطة بالاستجابة على أي فقرة أخرى في الاختبار.

3- الاطرادية (Monotonicity): وتعني أنه بازدياد قيمة القدرة θ، تزداد احتمالية الإجابة الصحيحة على الفقرة، أو تبقى ثابتة ضمن مستويات القدرة المختلفة، التي يعبر عنها رياضيا كالآتي، إذا كانت $\theta_a < \theta_b$.

$P(x_i = 1/\theta = \theta_a) \le P(x_i = 1/\theta = \theta_b)$ لجميع الفقرات.

4- الاطرادية المضاعفة (Double Monotonicity): وهو الافتراض الأصعب وغير الضروري لتحقق النموذج، والمتضمن امتلاك دوال استجابة غير متقاطعة لفقرات الاختبار التي تشكل التدريج.

تقسم النماذج اللابارامترية إلى قسمين رئيسين، هما:

1- **نموذج التجانس الاطــرادي** (MHM) (Monotone Homogeneous Model) وهــو نموذج استجابة الفقرة اللابارامتري والمعروف بنموذج موكن، ويستخدم لتحليل التــدريج للاستجابات الثنائية، وقد وصف كل من مولينار وسيجتسما (Sijtsma, 1998) تحليــل التدريج لموكن على أنه نسخة معدلة احتمالية لتحليل التدريج لجتمان، حيث اقترح موكن نموذج التجانس الاطرادي في عام (1971) واشترط أحادية البعد للمقياس، بالإضافة إلى أن تكون دالة استجابة الفقرة غير متناقصة (الاطرادية) كدالة للقدرة، إذ يختلــف بــشكل أساسي عن النموذجين ثنائي وثلاثي البارامتر في أن دالة استجابة الفقرة ليس بالضرورة أن تأخذ شكلا لوجستيا مما يجعل نموذج موكن أقل تقييدا للبيانات التجريبية من النمــاذج اللوجستية، كما يتيح هذا النموذج إمكانية ترتيب الأفراد تبعا لمستوى القــدرة باســتخدام الدرجة الكلية.

2- **نموذج الاطراد المضاعف** (DMM: Double Monotonicity Model)

هو نموذج موكن الثاني، الذي يفترض جميع افتراضات نموذج التجــانس الاطــرادي، بالإضافة إلى افتراض عدم تقاطع دوال الاستجابة لفقرات الاختبار، ولكن يسمح لها بالتمــاس فــي المنــاطق المتطرفــة، ممــا يجعــل منــه نموذجــا صــعب التحقيــق (Sijtsma& Molenaar,2002).

معامل التدريج (Scalability Coefficient)

يعد نموذج جتمان أساس معاملات التدريج الضروري توفرها (H, Hi, Hij) حيث عمــل موكن على استخدام معامل لويفنجر (H) في عام 1971 لإيجاد مجموعة جديدة من المعاملات التي تشير إلى التجانس الاطرادي وتحققه في المقياس، حيث يأخذ هذا المعامل بالاعتبار كــل

زوج من أزواج الفقرات داخل الاختبار وكل فقرة تبعا لباقي فقرات الاختبار بالإضافة إلى تدريج المجموعة كاملة من الفقرات (Linden & Hambleton, 1997).

قابلية التدريج (Scalability)

يشير موكن ولويس (Mokken & Lewis, 1982) إلى أن الهدف الأساسي من نماذج التدريج عامة هو تزويد الفاحص بطريقة لقياس جودة مجموعة من الفقرات في مطابقة تدريج معين، وقد عرّفا التدريج على أنه مجموعة من الفقرات التي ترتبط مع بعضها إيجابيا بحيث يكون معامل التدريج (Hi) أكبر من أو يساوي قيمة ثابتة موجبة (c)، حيث (0<c<1)، حيث (c) ثابت تعريف التدريج. في حال تحقق التعريف السابق على مجموعة الفقرات يطلق عليها أنها فقرات تتبع تدريج موكن.

دقة تقدير معالم الفقرة والقدرة

إن الخطوة الأساسية والأهم في تطبيق نظرية استجابة الفقرة (نظرية الفقرة) هي تقدير معالم نموذج نظرية الاستجابة للفقرة، التي تحدد خصائص هذا النموذج، كما أن النجاح في عملية التقدير يتوقف على توفير إجراءات مناسبة لتقدير هذه البارامترات، سواء في النماذج البارامترية أو اللابارامترية لنظرية استجابة الفقرة.

وتعتمد الدقة في تقدير البارامترات التي يتم الحصول عليها على مبدأين أساسيين، هما: عدم التحيز في التقدير(unbiased)، ومعدل مربعات الأخطاء (Expected Mean Squared Error: EMSE)، حيث يشير التحيز (Bias) إلى توقع الفرق بين تقدير المعلم، وقيمة المعلم الفعلية، ولا يعني تحقق هذا المبدأ أن التقدير دقيق تماما، وذلك لإمكانية وجود تقديرات عدة غير متحيزة، لهذا فإن مبدأ معدل مربعات الأخطاء للتقدير الذي يقاس بواسطة

تباين القيمة المقدرة يعطي أسلوبا لاختيار تقدير جيد للبـــارامترات (Casella & Berger, 1986).

وهناك خصائص أخرى مرغوب فيها فــي التقــديرات الإحــصائية، منهــا الاتــساق (Consistency) الذي يعني أنه كلما ازداد حجم العينة، فــإن قيمــة المقــدر (Estimator) تقترب من المعلم باحتمالية كبيرة.

وقد رأى وانق و فيزبول (Wang & Vispole, 1998) أن هناك ما لا يقل عن ثلاث حالات يمكن أن يكون التحيز هو الأكثر أهمية من الخطأ المعياري، وهي في مقارنة أوســاط المجموعات، وفي استرجاع تقديرات القدرة من الاختبارات المختلفة على التدريج نفسه، وفي تحديد الإتقان أو عدمه. وبالتالي تعد الطريقة التي تعطي تحيزا أقل هي الطريقة المفضلة.

وتشير الدقة إلى المدى الذي يتوافق فيه القرار المسند على درجات الاختبار مع القرار الذي يمكن اتخاذه فيما لو كانت الدرجات لا تتضمن أية أخطاء قياس، وبالتالي فإن الدقة لا بد من تقديرها على اعتبار أن الاختبار الذي لا يتضمن أخطاء غيــر موجــود (Crocker & Algina, 1986).

وفي نماذج نظرية استجابة الفقرة (البارامترية واللابارامترية) يعتمد احتمال الإجابــة الصحيحة على بارامتر القدرة (θ)، وعلى معالم الفقرة ذات العلاقة، وهي: بارامتر الصعوبة (b)، وباراميتر التمييز (a)، وباراميتر التخمين (c)، حيث إن جميع هذه البــارامترات غيــر معروفة، ولذلك فإن مهمة التقدير هي تحديد قيمة (θ) لكل مفحوص بالإضافة إلــى معرفــة معالم الفقرة من خلال إجابات المفحوصين على فقرات الاختبار. وهناك طرق عــدة لتقــدير معالم الفقرة والقدرة سواء في النماذج البارامترية أو اللابارامترية، وهــذه الطــرق تــستخدم أساليب التحليل العددي (Numerical Analysis) من خلال برامج يمكن تطبيقها باســتخدام

الحاسب الإلكتروني، وقد أشار ثيسن ووينر (Thisen & Wainer, 1982) إلى أن تحديد كمية الخطأ في تقدير البارامترات ليست عملية سهلة، وذلك بسبب عدم وجود صيغة رياضية محددة تعطي الخطأ في تقدير المعلم كدالة لحجم العينة وعدد الفقرات، ففي أبسط الصيغ الرياضية نحتاج إلى استخدام طرق التكامل العددي (Numerical Integration) لإيجاد الخطأ المعياري لتقدير معالم الدوال اللوجستية.

وهناك طرق عدة للقيام بتقدير معالم نماذج نظرية استجابة الفقرة البارامترية واللابارامترية، ولكن السؤال الذي يبقى قائما الذي يفترض أن تجيب عليه أي نظرية قياس (أو نموذج قياس) هو: ما هي دقة تقدير معالم الفقرة والقدرة؟

وقد بين لورد (Lord,1980) وجود وسائل عدة للكشف عن دقة تقدير البارامترات وجودة الاختبارات، من أبرزها:

- محك معدل مربعات الأخطاء (EMSE أو الخطأ المعياري في التقدير (S.E.E)
- محك الكفاءة النسبية للاختبار (RE: Relative Efficiency)، ودالة معلومات الاختبار (TIF: Test Information Function). هذا وسيتم في هذه الدراسة استخدام هذه المؤشرات لتحديد دقة تقدير البارامترات في النماذج البارامترية واللابارامتري لنظرية استجابة الفقرة.

أساليب تقدير معالم الفقرة والقدرة في النماذج البارامترية لنظرية استجابة الفقرة

هناك أسلوبان لتقدير معالم الفقرات والأفراد، الأول: يعتمد على تقديرات الأرجحية العظمى (Maximum Likelihood)، بينما الأسلوب الثاني: يعتمد على نظرية بيز (Bayes Theory).

وهناك ثلاث طرق رئيسية تعتمد عليها طريقة الأرجحية العظمى (ML) لتقدير معـــالم فقرات النماذج أحادية البعد (De Gruijter, Van Der & Kamp; 2005; Hambleton, 1989) وتختلف هذه الطرق في تعاملها مع مستويات القدرة غير المعلومة، وهي:

- تقديرات طريقة الأرجحية العظمـــى المـــشتركة (Joint Maximum Likelihood Estimation)، حيث يمكن تطبيق هذه الطريقة فـــي جميـــع النمـــاذج اللوجـــستية الأحادية والثنائية والثلاثية البارامترات، ويتم وفقا لهذه الطريقة تقدير معالم الفقرة، والقدرة في أن واحد معا. إذ تعمل هذه الطريقة على نمذجة احتمالات الاســتجابات باستخدام تقديرات مبدئية لمستويات قدرة الأفراد.

- تقديرات الأرجحية العظمـــى الـــشرطية (Conditional Maximum Likelihood Estimation)، وتقوم هذه الطريقة بفصل البارامترات الإحصائية للمفحوصين في أثناء عملية التدريج، وتطبق على النموذج اللوجستي الأحادي البارامتر، حيث يكون اقتران الاحتمالية (Likelihood) مشروطا بعدد الإجابات الصحيحة للأفراد علـــى الفقرات الاختبارية، لذلك يمكن الحصول على تقديرات الأرجحية العظمى الشرطية لمعالم صعوبة الفقرات بغض النظر عن بارامتر القدرة.

- تقـــديرات الأرجحيـــة العظمـــى الهامـــشية (Marginal Maximum Likelihood Estimation)، ويمكن تطبيق هذه الطريقة على النماذج اللوجستية الأحادية والثنائية والثلاثية المعلم، حيث يتم وفقا لهذه الطريقة إيجاد قيمة (θ) التي تجعل قيمتها (θ) أكبر ما يمكن من خلال معادلة رياضية، عن طريق إيجاد المشتقة الأولـــى لتلـــك المعادلة ومساواتها بالصفر، حيث يتم وفقا لهذه الطريقة إيجاد اقتـــران الاحتماليـــة الهامشي (Marginal Likelihood Function) لمعالم الفقرة، من خـــلال تكامـــل

اقتران الكثافة (Density Likelihood Function)على معالم القدرة ثم إيجاد تقديرات معالم الفقرات من خلال المعادلة الرياضية الآتية:

$$L(a,b,c) = \prod \int_{\infty}^{\infty} g(\theta_a) L(\theta_a; a,b,c) d\theta_a$$

أما أساليب تقدير بارامتر القدرة للأفراد، فيتم تقديرها بأسلوبين: يعتمد الأسلوب الأول على تقديرات الأرجحية العظمى (ML)، بينما يعتمد الأسلوب الثاني على تقديرات منهج بييز (Bayesian Estimation). ويعتمد برنامج (Bilog–Mg 3) الذي اعتمد في تحليل البيانات لهذه الدراسة (وفق النموذج البارامتري)، في تقدير معالم الفقرات طريقة الأرجحية العظمى الهامشية (Zimowski, Muraki, Mislevy & Bock, 2003). وتتميز هذه الطريقة بأنه يمكن استخدامها في تقدير معالم جميع النماذج أحادية البعد (البارامترية)، كما تتميز بالفعالية (Efficiency)، سواء كانت فقرات الاختبار قليلة أو كثيرة. والقيم التقديرية للأخطاء المعيارية الناتجة تتميز بالدقة، ويمكننا الحصول على قيم تقديرية لمعالم الأفراد الذين أجابوا إجابة صحيحة أو خطأ عن جميع الفقرات، وبالتالي لا يوجد فقدان للمعلومات عائد لحذف استجابة بعض أفراد الدراسة، إضافة إلى أنها تعطي تقديرات للعلامة الكلية، وأن التقديرات الناتجة تكون متسقة وتقترب من القيم الحقيقية كلما زاد حجم العينة (علام، 2005)، وتفترض هذه الطريقة التوزيع الطبيعي لتقديرات القدرة للأفراد، وتتضمن طريقتين لحل معادلة الأرجحية العظمى الهامشية: الطريقة الأولى: وتتضمن مرحلتين أحداهما: مرحلة التوقع (Expectation) والأخرى مرحلة التعظيم (Maximization)، ويرمز لها بالرمز (EM)، وهذا البرنامج يعتمد على تكرار الخطوات. أما الطريقة الثانية: تتضمن أسلوب نيوتن–جاوس (Newton–Gauss)، وهذه الطريقة تقوم بالبحث بطريقة متكررة، حيث يتم فيها تحسين القيم

التقديرية لهذه البارامترات بالتتابع، بحيث تحقق القيم النهائية محك التقارب (Zimowski, Muraki, Mislevy & Bock, 2003).

كما يعتمد برنامج (Bilog-Mg 3) طريقة الأرجحية العظمى (ML) في تقديرات القدرة للأفراد، ويتم تقدير القدرة للأفراد عن طريق الأرجحية العظمى (ML)، عندما تكون معالم الفقرات معروفة عن طريق الأرجحية العظمى الهامشية. ويتم تقدير معالم قدرة الأفراد (θ) في هذه الحالة عن طريق تعظيم اقتران دالة الأرجحية والتي تعطى بالعلاقة الآتية:

$$LogL_i(\theta) = \sum_{j=1}^{n} \{x_{ij} Log_e pj(\theta) + (1 - x_{ij}) Log_e [1 - pj(\theta)]\}$$

حيث إن" $p_j(\theta)$ احتمالية أن يجيب المفحوص ذو القدرة (θ) عن الفقرة (i) إجابة صحيحة.

i: رقم الفقرة، n: عدد فقرات الاختبار.

ومن ثم إيجاد تقديرات القدرة للأفراد (θ) التي تمثل أعلى ترجيح عن طريق إيجاد المشتقة الأولى لاقتران الترجيح ومساواته بالصفر والمعطى بالعلاقة الآتية:

$$\frac{\partial dLofLi(\theta)}{\partial \theta} = \sum_{j=1}^{n} \frac{xij - pj(\theta)}{pj(\theta)[1 - pj(\theta]} \frac{\partial pj(\theta)}{\partial \theta} = 0$$

وبعد ذلك، يتم القيام بالتدوير المتعاقب (Iteration) حسب طريقة نيوتن رافسون (Newton-Raphson Method)، من أجل الحصول على تقديرات ثابتة لمعلم القدرة من خلال معادلة فشر (Fisher-Scoring Solution). والخطأ المعياري المحسوب وفق هذه الطريقة يتم من خلال حساب الجذر التربيعي لمعكوس دالة المعلومات (Zimowski et al.,2003).

أما اقتران الاحتمالية الهامشي لدرجات المفحوصين للمتجه X فتعطى بالعلاقة الاتية:

$$fx_{i\setminus\delta}(X \setminus \delta) = \prod f_{xij/\delta}(X_i \setminus \delta)$$

وتعمل الأرجحية العظمى الهامشية على إيجاد تقديرات معالم الفقرات δ وذلك عن طريق تعظيم اقتران الأرجحية العظمى في المعادلة السابقة، كما في المعادلة الآتية:

$$L_1 = Logf_{x\backslash\delta}(x \backslash \delta)$$

وتعمل الأرجحية العظمى بشكل جيد في إيجاد تقديرات أكثر دقة لمعالم الفقرات، إذا كان حجم العينة كبيرا، أما إذا كان حجم العينة صغيراً، فإن التقديرات تكون أقل دقة، وخاصة بارامتر التخمين الخاص بالنموذج الثلاثي البارامتر (Birinbaum, 1986; Gao & Chen, 2005)

دقة القياس في نماذج استجابة الفقرة البارامترية

بعد الانتهاء من عملية تقدير معالم الفقرات والقدرات للأفراد، فإنه يجب التحقق من دقة هذه التقديرات، فقد بين لورد (Lord, 1980) وجود معايير عدة للتحقق من دقة التقدير، ومن أبرزها: دالة معلومات الفقرة والاختبار (Information Function)، والخطأ المعياري في التقدير (Standard Error of Estimation) (S.E.E)، والفعالية النسبية للاختبار (Relative Efficiency).

دالة معلومات الفقرة

لكل فقرة من فقرات الاختبار هناك ما يسمى دالة معلومات الفقرة (IIF)، وهو عبارة عن اقتران يبين مدى مساهمة الفقرة في تحديد القدرة. وبشكل عام، فإن الفقرات ذات التمييز العالي تسهم بقوة أكبر في تأكيد دقة القياس من تلك ذات التمييز المتدني، حيث تعطي الفقرة أفضل مساهمة لها في تأكيد دقة القياس حول قيمة صعوبتها (b) على متصل القدرة، فإذا ما كان المنحنى مزاحا باتجاه اليمين، فهذا يعني أن الفقرة صعبة، وإذا كان ارتفاعه كبيرا فهذا يعني أن تمييز الفقرة عالٍ.

وإن دالة معلومات الفقرة (IIF) تقوم بدور كبير في نظرية الاستجابة للفقرة (نظرية الفقرة)، حيث يمكن من خلالها تحديد الخطأ المعياري في التقدير، فعندما يتم استخراج تقدير الأرجحية العظمى لمعلمة القدرة (أو أي أسلوب تقدير)، فإن تباين الخطأ في تقدير القدرة يساوي معكوس دالة المعلومات وبالمثل عند ايجاد تقديرات لمعالم الفقرة، فإن مصفوفة التباين والتباين المشترك (variance covariance matrix) للتقديرات تساوي معكوس مصفوفة المعلومات (Information matrix) لتقديرات معالم الفقرة، وتعتبر مصفوفة التباين المشترك لتقديرات معالم الفقرة على قدر من الأهمية عندما يتم مقارنة معالم الفقرة لمجموعتين من المفحوصين (Hambleton & Swaminathan, 1991).

وتختلف معادلة منحنى معلومات الفقرة حسب النماذج اللوجستية، ففي النموذج اللوغاريتمي أحادي البارامتر (نموذج راش)، فإن معادلة منحنى معلومات الفقرة تعطى بالعلاقة الآتية:

$$I_i(\theta) = P_i(\theta)\, Q_i(\theta)$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

أما في النموذج اللوغاريتمي ثنائي البارامتر تعطى بالعلاقة الآتية:

$$I_i(\theta) = a_i^2\, P_i(\theta)\, Q_i(\theta)$$

أما في النموذج اللوغاريتمي ذي البارامترات الثلاث، فإن معادلة منحنى معلومات الفقرة تعطى بالعلاقة الآتية:

$$I_i(\theta) = a_i^2 \left[\frac{Q_i(\theta)}{P_i(\theta)}\right] \left[\frac{(P_i(\theta) - c)^2}{1 \quad -c}\right]$$

وقد أشار بيكر (Baker, 2001) إلى أن منحنى معلومات الاختبار يقدم فوائد كثيرة، إذ يقدم معلومات عن دقة الاختبار في تقدير قدرات الأفراد، وأشار إلى أن الوضع المثالي يتطلب

أن يكون هذا المنحنى أفقياً، إلا أن هذا الوضع المثالي لا يتناسب مع أغراض بعض الاختبارات، فمثلا في اختبارات الاختيار التي نهدف من خلالها إلى اختيار مجموعة من الأفراد أو الطلاب دون آخرين، فإن الهدف في هذه الحالة فصل الطلاب إلى مجموعتين مختلفتين، وبناء عليه يكون المنحنى المفضل هو المنحنى الذي تقع قمته عند نقطة القطع.

ويذكر بيكر أن تحديد دقة القياس في النماذج البارامترية المختلفة يتم من خلال دالة معلومات الفقرة، التي تزودنا بالمعلومات التي تقدمها الفقرة عند كل مستوى من مستويات القدرة والمعطاة بالمعادلة الآتية:

$$I_i(\theta) = \frac{(p_i(\theta))'}{p_i(\theta)Q_i(\theta)}$$

حيث إن:

$$Q_i(\theta) = 1 - Pi(\theta)$$

($(p_i(\theta))'$): هي المشتقة الأولى لاحتمالية الإجابة الصحيحة على الفقرة (i) تبعاً للمتغير القدرة (θ)

ويضيف بيكر أنه وبعد استخراج دوال معلومات جميع فقرات الاختبار يتم استخراج دالة معلومات الاختبار التي تحدد حجم المعلومات التي يزودنا بها الاختبار ككل عند كل مستوى من مستويات القدرة، التي هي حاصل جمع دوال معلومات فقرات الاختبار ككل والمعطاة بالمعادلة الآتية:

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta)$$

الخطأ المعياري في التقدير

يعرف الخطأ المعياري في القياس حسب النظرية الكلاسيكية (نظرية الاختبار) بالصورة الرياضية الآتية:

$$SE_m = \sigma_x \sqrt{1 - P_{xx}}$$

حيث إن (ρxx) هو معامل الثبات، σx هو الانحراف المعياري للدرجات الخام، ولذلك فإن محدودية مفهوم النظرية الكلاسيكية لدقة القياس تم تجاوزه وفقا لمبادئ نظرية الفقرة (نظرية الاستجابة للفقرة) عبر دالة معلومات الفقرة، وتستخدم دالة معلومات الفقرة في تحديد الدقة التي تقدر بها قدرات المفحوصين عند المستويات المختلفة.

والخطأ المعياري في القياس حسب نظرية الاختبار يناظر الخطأ المعياري في التقدير حسب نظرية الفقرة في القياس، وحيث إن مربع الخطأ المعياري في التقدير يرتبط بالثبات (Reliability)، فإن ذلك يشير إلى أنه كلما قل مقدار الخطأ المعياري، فإن ذلك يؤدي إلى زيادة الثبات (الدقة) (Reeve, 2004).

ولذلك فإن زيادة عدد الفقرات يعطي خطأ معياريا صغير (θ) S.E، ونقصان قيمة الخطأ المعياري في تقدير القدرة عند مستوى القدرة يؤدي إلى زيادة كمية المعلومات للاختبار وفق العلاقة الآتية:

$$SEE = 1/\sqrt{\text{Information}}$$

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta)$$

$$I(\theta) = (1/S.E.(\theta))^2$$

ونقصان قيمة الخطأ المعياري في تقدير القدرة يؤدي إلى زيادة قيمة معامل الثبات شأنه في ذلك شأن النظرية الكلاسيكية. وما يميز هذه الطريقة عن الطرق الكلاسيكية أن تقدير

الثبات في النظرية الكلاسيكية مرتبط بالعينة، وهذه سلبية من سلبيات تقدير الثبات، كما أنها تؤدي لتقدير عام للأخطاء الفردية في علامات الاختبار، وهو ما يسمى بالخطأ المعياري في القياس، بينما نظرية استجابة الفقرة تزودنا بتقدير للخطأ المعياري للقياس عند كل مستوى من مستويات القدرة، ويمكن باستخدامها تحديد مدى مساهمة كل فقرة في دقة الاختبار (Hambelton & Swaminathan, 1985).

الفعالية النسبية للاختبار

يوجد تطبيقات متعددة في القياس النفسي والتربوي تتطلب مقارنة الفعالية النسبية (Relative Efficiency) لاختبارين في قياسهما قدرة معينة عند نقط مختلفة على متصل القدرة، إذ ربما نحتاج إلى مقارنة دوال معلومات اختبارات تشتمل على فقرات مختلفة مستمدة من نطاق سلوكي معين. ويمكن إجراء هذه المقارنات بين دوال المعلومات بحساب الفعالية النسبية لأحد الاختبارات مقارنة باختبار آخر.

ويمكن التعبير عن ذلك بالصيغة الرياضية التالية لدالة الفعالية النسبية (Relative Efficiency Function(REF)):

$$RE(\theta) = \frac{I_a(\theta)}{I_b(\theta)} = \left(\frac{SE_b(\theta)}{SE_a(\theta)}\right)^2$$

حيث إن $RE(\theta)$ تشير إلى الفعالية النسبية، $I_a(\theta)$: دالة معلومات الاختبار a، $I_b(\theta)$: دالة معلومات الاختبار (b) (Hambleton & Jones,1993).

ودرجات الاختبارين يمكن أن تكون الدرجات على اختبارين مختلفين عند مستوى القدرة نفسه، أو للاختبار نفسه بطريقتين مختلفتين. وتختلف الفعالية النسبية لاختبارين b , a باختلاف مستوى القدرة.

ويرى تريفسان وسايس وميشيل (Trevisan, Sax & Michael, 1991) أن التنبؤ بدرجات الأفراد على الاختبار يكون أكثر دقة إذا ربطنا ذلك بمستويات القدرة، حيث وجـدوا أيضا أن الاتجاه المفضل لعدد البدائل هو أربعة بدائل، ثم ثلاثة بدائل، ثم خمسة بدائل للأفراد ذوي القدرة المنخفضة، أما بالنسبة للطلبة ذوي القدرة المرتفعة وذوي القدرة المتوسطة، فـلا يوجد تفضيل لعدد البدائل، وكان العدد المثالي للبدائل عبر مجموعات القدرة هو أربعة بدائل، وذلك باستخدام النموذج اللوجستي ثلاثي البارامتر والمستخدم في هذه الدراسة.

ومن الجدير بالذكر أن دالة الفعالية النسبية يتم حسابها عند كل نقطة من نقاط متصل القدرة كما هو الحال في دالة المعلومات. غير أن دالة المعلومـات لا يمكـن تفـسيرها فـي عبارات مطلقة ما لم يتم تعريف تدريج قياس القدرة (θ–metric)، بينما لا يؤثر ذلك في دالة الفعالية النسبية، وهذا يضفي أهمية على هذه الدالة الأخيرة ويثري استخداماتها في المقارنـة بين الاختبارات وغيرها من التطبيقات السيكومترية (علام، 2005).

أساليب تقدير معالم الفقرة والقدرة في النماذج اللابارامترية

هناك أساليب عدة لتقدير معالم الفقرة والقدرة حسب نماذج نظرية اسـتجابة الفقـرة اللابارامترية، وتختلف هذه الأساليب عن نظيرتها البارامترية في أنها لا تفتـرض أي شـكل حسابي سابق، وتقدر منحنى خصائص الفقرة (ICC)، ثم تجد البارامترات من هذا المنحنـى، وذلك أنها تفترض وقوع البيانات على مستوى القياس الرتبي، ومن هذه الأساليب ما يأتي:

الانحدار اللابارامتري (Non–Parametric Regression)

أشار دوجلس (Douglas, 1997) أن تقدير دالة استجابة الفقرة ضمن تحليل الانحدار اللابارامتري يتم دون أي افتراضات فيما يخص شكل هذه الدالة (لوجستية الشكل) كما في النماذج البارامترية، حيث يوجد على الأقل طريقتين لتقدير شكل هذه الدالة، وهما:

1. Kernel Smoothing (KS).
2. Isotonic Regression Estimation.

ويضيف أن الأفضلية التطبيقية تعود إلى طريقة (KS)؛ وذلك لوجود برنامج حاسوبي مجاني (TESTGRAF)، الذي يمكن من خلاله تقدير دالة Pi(θ)، حيث يقوم (KS) بحساب Pi(θq)، الذي يدل على احتمالية الإجابة الصحيحة على الفقرة i عند مستوى (q) معين من القدرة من خلال المعادلة الآتية:

$$p_i(\theta_q) = \sum_{a=1}^{N} w_{aq} y_{ima}.$$

حيث إن

$$w_{aq} = \frac{k\left[\frac{\theta_a - \theta_q}{h}\right] y_i}{\sum_{b=1}^{N} k\left[\frac{\theta_a - \theta_q}{h}\right]}$$

- w_{aq} : متجهة القدرة للمفحوص a عند مستوى قدرة q، الذي يتم تقديره تبعاً لرتبة المفحوص a مع رتب باقي المفحوصين (b , c , ... , N).
- y_{ima} : متجه خيار الفقرة الثنائي بطول يساوي N (عدد المفحوصين الكلي) ، والذي يأخذ القيمة 1 في حال اختار المفحوص a الخيار m.
- K : دالة كيرنل التي يمكن تقديرها بطرق عدة باستخدام برنامج (TESTGRAF).
- h : بارامتر التهذيب (Smoothing Parameter)، وهو يعتمد بشكل أساسي على عدد المفحوصين، ويساوي (1.1N) في برنامج TESTGRAF (RAMSAY, 2000).

وتم في هذه الدراسة استخدام هذه الطريقة (Kernel Smoothing) لتقدير معالم الفقرة والقدرة لنماذج نظرية استجابة الفقرة اللابارامترية، وهي:

The Nonparametric Gaussian Kernel Smoothing Model

وتصف مناهج (IRT) اللابارامترية احتمالات النجاح في فقرات الاختبار كدوال رتبية للقدرة دون افتراض أي شكل حسابي سابق. وحسب فاندير لندن وهامبلتون (Vander Linden & Hambleton ,1997)، يمكن الافتراض أن هذه الدوال أقرب لدوال الاستجابات الحقيقية من تلك التي تعطيها النماذج البارامترية؛ لأنها تعتمد على افتراضات أقل حول النموذج الحسابي.

وقديما، خدمت النماذج اللابارامترية كبدائل للنماذج البارامترية عندما لا تكون الظروف مناسبة لاستخدام نماذج بارامترية معينة، مثل تدريج أو قياس الاتجاهات (Mokken,1997). إلا أن رامسي (Ramsay, 1997) ينادي بان التقدير اللابارامتري لمنحنى الاستجابة قد يصبح نموذج الاختيار، إلا إذا كان هناك أسباب كبيرة وقوية لاستخدام نموذج بارامتري معين، وتقدر النماذج البارامترية هذه المنحنيات عن طريق تقدير البارامترات في حين تقدرها النماذج اللابارامترية مباشرة من المقادير الملاحظة.

ويضم منهج (Kernel Smoothing) للتحليل اللابارامتري تقديرات القدرة وتحليل الفقرات في برنامج (TESTGRAF)، ويتم حساب قيم منحنى خصائص البديل بالطريقة الآتية (Ramsay,1995):

أولا: يتم الحصول على العلامة الكلية (Xa) لكل متقدم a حيث a =(1، 2، 3....N) وXa هو عدد الإجابات الصحيحة، ويتم إعطاء المتقدمين ذوي العلامات الكلية المتساوية ترتيبا رتبيا بشكل عشوائي.

ثانيا: يتم إعطاء المتقدمين علامات ربعية (كمية) للتوزيع الطبيعي المعياري اعتمادا على رتبهم (Za)، بحيث تكون المساحة تحت دالة الكثافة الطبيعية المعيارية إلى يسار (Za) مساوية a/(N + 1).

ثالثا: يتم تسجيل قيم المؤشر (Yjka =1) عندما يقوم المتقدم (a) باختيار البديل k، وغير ذلك فهو صفر للمتقدم على الرتبة a للبديل k للفقرة j.

وأخيرا تم تقدير احتمال اختيار البديل k للفقرة (j)،[Pjk(θ)] عن طريق تلطيف العلاقة بين متجه المؤشر [Yjka] من الرتبة N ومتجه الكفاءة (Zas) ل(N) من المتقدمين. ويعتبر التلطيف نوعا من المعدلات التي تكون فيها ((Pjk(θ)) عند أي مستوى ل (θ) هو المتوسط الموزون ل (Yjka) للمتقدمين ذوي مستويات القدرة القريبة من (θ) وتظهر معادلة التلطيف التي أعطاها رامسي (1995) بالشكل الآتي:

$$\hat{P}_{jk}(\theta_q) = \frac{\sum_{a=1}^{N} K\left[(\theta_a - \theta_q)/h\right] y_{ajk}}{\sum_{a=1}^{N} k\left[(\theta_a - \theta_q)/h\right]}$$

وتم في هذه المعادلة السابقة استبدال (θ_a) بالتقدير الأولي (Za). وقد مثلت (θ_a) نقاط التقدير، وعادة هي النقاط المتوسطة لفترات متساوية المسافات على مقياس القدرة. K هي الدالة القوسية K(u)=exp(−u2/2) حيث u =(θ_a − qθ)/h حيث h بارامتر التبسيط الذي يتحكم بحجم تحيز الاستبدال (bias)، ومع زيادة قيمة (h) يزداد التحيز في حين يتراجع تباين العينات. وبشكل طبيعي، يستخدم برنامج Testgraf ($h = 1.1N^{-1/5}$) بحيث يتم تقليل مربع متوسط الخطأ، ومجموع التحيز المربع، وتباين العينات، وبحجم عينة أكبر من (1) تكمن قيمة (h) الطبيعية في مدى صفر إلى 1. وحيث إن h دالة مقلوبة لـــ (N)، ترافقت قيمة (h) الأكبر مع (N) أصغر أو بشكل متكافئ، هناك حاجة أكثر للتبسيط بالعينات الصغيرة، حيث يكون مقام المعادلة وببساطة عامل تدريج لاقتصار دالة الأشكال على المدى من صفر إلى واحد.

ولتسريع عملية الحساب دون التضحية بالدقة، فقد تم الحصول على معدل قيم المؤشر ($\hat{P}\, jkr$) لقيم θ_a التي تقع بين مركزي الفقرات المتجاورة، وكذلك تم حساب المنطقة التي تقع تحت المنحنى الطبيعي المعياري بين مراكز هذه الفترات (ϕ_r). وبعد ذلك تم تبسيط هذه القيم كما بالمعادلة الآتية:

$$\hat{P}_{jk}(\theta_q) = \frac{\sum_{r=1}^{Q} \phi_r k\left[(\theta_q - \theta_r)/h\right]\hat{P}_{jkj}}{\sum_{r=1}^{Q} \phi_r k\left[(\theta_q - \theta_r)/h\right]}$$

وبعد ذلك تم تقدير تباين الخطأ لقيمة المنحنى المقدر بالمعادلة الآتية

$$\hat{S}(\theta_q) = \sum_{a=1}^{N} w_{jka}^2\ \hat{P}_{jk}(\theta_a)(1-\hat{P}_{jk}(\theta_a))$$

$$w_{jka} = \frac{K\left[(\theta_a - \theta_q)/h\right]}{\sum_{a=1}^{n} K\left[(\theta_a - \theta_q)/h\right]}$$

وتم تقدير تباين الخطأ لقيمة المنحنى المقدر $P_{jk}(\theta_a)$ بالمعادلة التي تليها مباشرة، وتم إعطاء الوزن المستخدم في تقدير تباين الخطأ بالمعادلة السابقة .(Ramsay, 1995)

وبعد أن يتم تقدير (OCC)، يمكن عندها تقدير قدرة المتقدمين للاختبار اعتماداً على هذه المنحنيات. ويمكن تدوير تقديرات القدرة الجديدة للوراء في العملية كأساس للترتيب وإعادة تقدير المنحنيات، ويمكن الاستمرار بهذه المحاولة حتى تتشابه تقديرات المنحنى والقدرة. وقد ذكر رامسي (Ramsay, 1995) أن محاولتين إلى ثلاث محاولات تكون كافية للوصول إلى التشابه.

دقة القياس في نماذج استجابة الفقرة اللابارامترية

يشير رامسي (Ramsey, 2000) إلى أنه يتم تقدير دقة القياس في النماذج اللابارامترية من خلال دالة معلومات الفقرة والاختبار كما هو الحال في النماذج البارامترية، حيث تعطى دالة معلومات الفقرة للفقرات الثنائية كما يأتي:

$$I_i(\theta)=\left[\frac{dp_i(\theta)}{d\theta}\right]^2/[p_i(\theta)(1-P_i(\theta))]$$

حيث إن $Pi(\theta)$: قيمة دالة خصائص الفقرة.

ويتم تحديد دقة القياس في النماذج اللابارامترية من خلال استخدام برنامج (TESTGRAF) برسم متوسط دالة معلومات الفقرة (Mean Item Information Function). وتستخرج دالة معلومات الاختبار بالاعتماد على دوال معلومات فقرات الاختبار كامل حيث تعطى بالمعادلة الآتية:

$$I(\theta)=\sum_{i=1}^{n}I_i(\theta)$$

ثم يستخرج متوسط دالة معلومات الاختبار (Average East Information Function)، حيث يقوم برنامج (TESTGRAF) بعمل مقارنات بين اختبارات مختلفة قد تحتوي أعداداً مختلفة من الفقرات.

التطبيقات البارامترية واللابارامترية في نظرية استجابة الفقرة

توضح نظرية استجابة الفقرة العلاقة بين قدرات المفحوصين واستجاباتهم على فقرات الاختبار بنماذج إحصائية (Crocker & Algina, 1986). وتنقسم هذه النماذج إلى نوعين رئيسين، نماذج استجابة الفقرة البارامترية (PIRT: Parametric Item Response Theory)حيث يكون شكل دالة استجابة الفقرة (IRF: Item Response Function)

محدداً، ونماذج استجابة الفقرة اللابارامترية (- Non Item Response Theory (NIRT) Parametric)، التي لا تحدد شكل دالة استجابة الفقرة.

وبالرغم من أن نماذج الاستجابة للفقرة تقدم مزايا تتعلق بتفسير علامات الاختبار، إلا أن هذه المزايا تتحقق من ناحية عملية فقط عند وجود اتساق بين النموذج المستخدم وبيانات الاختبار؛ بمعنى أن أهمية اختيار أحد نماذج الاستجابة للفقرة تبرز في قدرته على تفسير البيانات، وفي الدرجة التي تتطابق فيها البيانات مع افتراضات النموذج، الذي يخدم التطبيق العملي الذي يهتم فيه البحث مثل بناء اختبار لتصنيف الأفراد أو اختيار أفضل اختبار من بنك أسئلة من الفقرات المدرجة أو غيرها من التطبيقات (Hambleton & Swaminathan, 1985).

ويوجد اعتبارات عدة يجب أخذها بعين الاعتبار عند اختيار النموذج المناسب للبيانات سواء النماذج البارامترية أو النماذج اللابارامترية. فقد شاع استخدام نظرية استجابة الفقرة البارامترية من قبل الباحثين على الرغم من موانعها في تحليل البيانات ذات المستوى الرتبي، إلا أن مصداقية النتائج قد تكون موضع تساؤل عندما لا يتحقق فرض وقوع البيانات على مستوى القياس الفئوي، الأمر الذي تبرره نماذج استجابة الفقرة اللابارامترية والتي لا تضع قيود حول شكل دالة استجابة الفقرة، مما يثير التساؤل حول مدى مطابقة النوعين للبيانات التحصيلية التي يعتبرها البعض رتبية والبعض الآخر يعتبرها شبه فئوية، ومدى دقة النتائج التي تفرزها مثل هذه الاختبارات (Liang, 2010).

وتعتبر مطابقة النموذج للبيانات ذات المستوى الرتبي أحد الأمثلة على مشاكل المطابقة ومن ثم على دقة النتائج. حيث قام بعض الباحثين بتقديم بعض الاعتراضات على استخدام نماذج استجابة الفقرة البارامترية للبيانات الرتبية (والتي تهدف إلى ترتيب الأفراد على متصل

السمة)، حيث سمح البعض استخدامها في حال توفرت افتراضات النظرية لجميع البيانات الرتبية (Crocker & Algina, 1986)، في حين ركز البعض على التمييز بين المقاييس الرتبية ومقاييس الفترة على الإحصاءات البارامترية مع تطوير الحلول المناسبة، حيث أوصى البعض باستخدام النماذج اللابارامترية في تحليل البيانات الرتبية؛ لأن المقاييس التربوية نادرا ما تثبت أنها تتحلى بقوة خصائص مقاييس الفترة (Mokken & Lewis, 1982).

وقدم كل من جنكر وسيجستما (Junker & Sijtsma, 2001) ثلاثة أسباب تبين فائدة نماذج استجابة الفقرة اللابارامترية مقارنة مع نظيرتها البارامترية، وهي:

1. تقدم النماذج اللابارامترية فهما أعمق لما تقوم به نماذج استجابة الفقرة البارامترية.
2. تقدم النماذج اللابارامترية إطارا أكثر مرونة في حال ضعف المطابقة للنماذج البارامترية.
3. تقدم النماذج اللابارامترية طرقا وإجراءات سهلة للبيانات وتسمح باستخدام عينات أقل من الفقرات والأفراد (حجم عينة وطول اختبار اقل) عن نظيرتها البارامترية.

وإن كبر حجم عينة الأفراد والفقرات التي يتطلبها تطبيق النظرية الحديثة في القياس (Crocker and Algina ,1986) والدالة الرياضية المعقدة التي تفترضها لكل نموذج من نماذجها، أظهرت الحاجة إلى توفر برامج الحاسوب التي تساعد في تحليل بياناتها. ولقد ظهرت برامج حاسوبية كثيرة تستخدم في تحليل بيانات نماذج نظرية استجابة الفقرة، ففي عام 1985 صمم كل من مسلافي (Mislevy) وبوك (Bock) برنامج يسمى BILOG وهو يعتبر من البرامج القوية والفعالة التي تتسم بالمرونة والسرعة في تقدير معالم الفقرات والقدرة باستخدام الأرجحية القصوى الهامشية (MML: Marginal Maximum Likelihood)، وهذا البرنامج يستخدم لنماذج نظرية استجابة الفقرة البارامترية. وفي عام (1991) ادعى رامسي أن إجراءات تقدير برنامج (TESTGRAF) أسرع 500 مرة من البرامج الأخرى دون

خسارة في الكفاءة، وأنه يحتاج إلى حجم عينة وطول اختبار أقل من البرامج الأخرى، حيث إن هذا البرنامج يستخدم لنماذج نظرية استجابة الفقرة اللابارامترية، كما انتشرت البرامج التي تساعد الباحثين في مختلف العلوم من إنجاز الأبحاث والتوصل السريع للنتائج. ومن بين هذه البرامج، برنامج توليد البيانات في علم القياس والتقويم حيث تطورت برامج التوليد المتعلقة بنظرية استجابة الفقرة مع ظهور هذه النظرية في السبعينات ومن أجل تحديد خصائص وفعالية النماذج الرياضية وغيرها من الصيغ الرياضية التي تضمنتها هذه النظرية.

البرامج المستخدمة في الدراسة

أ. برنامج WinGen لتوليد بيانات الاختبار باختلاف متغيري (طول الاختبار، وحجم العينة):

يستخدم هذا البرنامج لتوليد القيم الحقيقية لمعلمة القدرة (θ)، ولمعالم الفقرات (a, b, c)، الذي يمكن من خلاله توليد بيانات أحادية البعد، ومتعددة الأبعاد، حيث يتم اختيار عدد المفحوصين، ونوع التوزيع لمعلم القدرة (طبيعي، وبيتا، ومنتظم)، وتحديد عدد الفقرات اللازمة وعدد الاستجابات ونوع النموذج (أحادي، وثنائي، وثلاثي) المعلمة، وتحديد التوزيع لمعالم الفقرات (طبيعي، وبيتا، ومنتظم)، وتوليد استجابات اعتماداً على معلمة القدرة ومعالم الفقرات، ويمكن تخزين ملفات تحتوي على قيم القدرات، ومعالم الفقرات، واستجابات المفحوصين، ومعاملات ارتباط الفقرات بالاختبار، ليتم معالجتها بالبرامج الأخرى التي تخدم نماذج نظرية استجابة الفقرة البارامترية واللابارامترية.

ب. برنامج (BILOG-MG) لتقدير قيم معالم الفقرات ومعلمة القدرة للنموذج الثلاثي البارامتري

تم استخدام هذا البرنامج في هذه الدراسة لتقدير معالم الفقرات (الصعوبة، والتمييز، والتخمين)، ومعلمة القدرة للأفراد (للنموذج الثلاثي المعلمة البارامتري)، بعد توليد البيانات من برنامج WinGen، للحصول على القيم التقديرية لمعالم الفقرة والقدرة حسب النموذج البارامتري لنظرية استجابة الفقرة.

ويوفر هذا البرنامج أوامر يمكن من خلالها استخدام النماذج اللوجستية الثلاثة: النموذج أحادي المعلم، والنموذج ثنائي المعلم، والنموذج ثلاثي المعلم. كما يعطي هذا البرنامج تقديرات لصعوبة الفقرة، وكذلك معلمي التمييز والتخمين حسب النموذج المستخدم، إضافة إلى تقديرات القدرة للمفحوصين، ويستخدم هذا البرنامج طريقة الأرجحية القصوى الهامشية (MML) في التقدير، ويتعامل هذا الأسلوب مع مستويات القدرة غير المعلومة، وذلك بالتعبير عن احتمالات أنماط الإجابات بتوقعات Expectation من توزيع مجتمع معين. فالبيانات الاختبارية ينظر إليها على أنها عينة عشوائية مستمدة من مجتمع. غير أنه لم تكن هناك إجراءات عملية لتقدير معالم الفقرات ومعلم القدرة للأفراد إلى أن توصل بوك وايتكن (Bock & Aitken, 1981) إلى برنامج حاسوبي يجري هذا التقدير في مرحلتين: إحداهما تسمى مرحلة التوقع Expectation، والأخرى تسمى مرحلة التعظيم Maximization. وهذا البرنامج يعتمد على تكرار الخطوات Iteration وهذا التكرار يؤدي إلى تحسين العدد المتوقع للإجابات الصحيحة، ومستويات القدرة. ويتميز هذا الأسلوب بميزات متعددة، لعل أهمها أنه يمكن استخدامه في تقدير معالم جميع النماذج أحادية البعد، كما يمكن الحصول على قيم

تقديرية لمعلم قدرة الأفراد الذين أجابوا إجابة صحيحة أو خطأ على جميع الفقرات (علام،2005).

ج. برنامج (TESTGRAF) لتقدير قيم معالم الفقرات ومعلمة القدرة للنموذج الثلاثي اللابارامتري

تم استخدام هذا البرنامج في هذه الدراسة لتقدير معالم الفقرات (الصعوبة، والتمييز، والتخمين)، ومعلم القدرة للأفراد (للنموذج الثلاثي المعلمة اللابارامتري)، بعد توليد البيانات من برنامج WinGen، للحصول على القيم التقديرية لمعالم الفقرات ومعلم القدرة للأفراد حسب النموذج اللابارامتري لنظرية استجابة الفقرة.

ويستخدم هذا البرنامج لنماذج نظرية استجابة الفقرة اللابارامترية، ويستخدم أسلوب تنعيم النواة (Kernel Smoothing) بدل أسلوب الأرجحية العظمى لتقدير معالم الفقرة والقدرة، ولا تقل كفاءته عن غيره من البرامج، ويستخدم للبيانات الثنائية (1, 0) ويحتاج حجم عينة وطول اختبار أقل من البرامج الأخرى، حيث إنه يحتاج (100) مفحوص و(20) فقرة لتقدير منحنى خصائص الفقرات (ICCs). ويتم تقدير منحنى خصائص الفقرة كما يأتي:

- يتم حساب علامة إجمالية (كلية) لكل مفحوص (j) عن طريق حساب النسبة المئوية للفقرات التي أجابها المفحوص إجابة صحيحة.
- يتم تصنيف المفحوصين وإجاباتهم على أساس علاماتهم الكلية (X_j).
- يتم إعطاء المفحوص المقدار (j) من التوزيع الطبيعي المعياري (Zj) حيث (Zj) هي المساحة أو المنطقة التي تقع دالة الشدة الطبيعية المعيارية تحتها إلى اليسار من (Zj)، التي تكون مساوية ل [1 + j / N].

ثم يتم تقدير العلامة Pjm(θ) عن طريق تبسيط أو صقل العلاقة بين قيم المتغير الثنائي (1 , 0) والمقادير الطبيعية المعيارية (Zj)، ويعتبر الصقل نوع من عمل التوسط، يتم فيه وزن احتمال الخيار Pimj عند كل مستوى قدرة (θ)، ومن ثم تتم عملية التقدير بطريقة (Kernel Smoothing) لحساب تقديرات معالم الفقرة والقدرة.

مشكلة الدراسة

يعد اختيار النموذج المناسب لتحليل البيانات الخطوة الأساسية في القياس النفسي والتربوي عموما، وعلى وجه الخصوص في نظرية استجابة الفقرة (نموذج بارامتري، أو نموذج لابارامتري)؛ لأن ذلك يلعب دورا كبيرا في مدى ملاءمة البيانات للنموذج المستخدم، وبالتالي يؤثر في دقة تقديرات النتائج وفهمها وتفسيرها. ويعتمد اختيار النموذج على نوع البيانات المتوفرة ومستوى القياس الذي تقع عليه هذه البيانات، والهدف الذي يسعى إليه الباحث من عملية القياس.

وبالرغم من الانتشار الواسع لنظرية الفقرة والميزات التي تمتاز بها، إلا أن هناك بعض المحددات لاستخدامها في تقدير معالم الفقرة ومعلمة القدرة، حيث إن نظرية الفقرة تحتاج إلى حجوم عينات وأطوال اختبار كبيرة نسبيا، إلا أن نماذج نظرية الفقرة اللابارامترية تحتاج إلى حجوم عينات وأطوال اختبار أقل من نظيرتها البارامترية، ولكن يبقى السؤال العام الذي يجب أن يجيب عنه أي نموذج تقدير، ما هي دقة تقدير هذه المعالم؟

ويحاول الفاحص في أي متغير يتعامل معه أن يحدد أرقاما أو رموزا ذات معنى تدل على سمة أو فئة في ذلك المتغير، أو أن يحدد مواقع الأفراد على تدريج معين بحيث يكون لفروق المواقع معنى يتناسب مع الفروق الحقيقية في مقدار ما يمتلكون من السمة موضوع القياس، سواء بمقارنة الموقع للفرد الواحد مع نقطة مرجعية يتم الاتفاق عليها، ويشار إليها

بالصفر الافتراضي (Arbitrary Zero)، أو أنها ذات معنى أو مدلول ثابت، وتعني لأي فرد ما تعنيه لأي فرد آخر، ويشار إليها بالصفر المطلق (Absolute Zero). وهذا يعني أن المقياس (Scale) يعرف من خلال الغرض، ويتلخص في تحديد مواقع الأفراد حسب نوع السمة أو حسب درجة امتلاكهم لها. كما أشار كيرلنجر (Kerlinger, 1973) إلى أن القياس ليس أكثر من لعبة ذات قواعد محددة، يسعى فيها اللاعب للوصول إلى أفضل تماثل (Isomorphism) بين القياس (Measurement) والحقيقة (Reality). وكما هو معروف لا نستطيع أن نعرف هذه الحقيقة، ولكن نعمل على تقدير لهذه الدرجة أو الترتيب والمهم في هذه العملية (التقدير) هو الوصول إلى التوافق التام في المواقع (Rank Isomorphism)، وليس القيم (Value Isomorphism).

وبذلك لا يوجد شك بإمكانية نجاح عملية القياس على مستوى القياس الاسمي أو الرتبي، وذلك أن معظم السمات النفسية والتربوية لا تقع على مقياس النسبة (لعدم توافر الصفر المطلق فيها)، ولا تقع على مقياس فئوي لعدم تساوي الفترات بين القيم (الفرق بين الدرجتين 95 و90 لا يساوي الفرق بين الدرجتين 55 و50). وبذلك فإن القياس النفسي والتربوي يقع على مقياس أرقي من المقياس الرتبي وأقل من الفئوي؛ أي شبه فئوي (Quasi Interval) (عودة، 2010).

ولكن المهتمين في القياس النفسي والتربوي يحللون نتائج القياس بافتراض أنها واقعة على مستوى القياس الفئوي؛ لأن معظم المعالجات الإحصائية اللازمة لتفسير النتائج وتحليلها مبنية على الأقل على مقياس فئوي، ولأن هذه البيانات تمكن المختص من تحويلها تحويلا خطيا، مع العلم أن انتهاك هذا الافتراض (وقوع البيانات على مقياس فئوي) قد يؤدي إلى تشويه النتائج وصعوبة في التفسير.

من هنا جاءت مبررات هذه الدراسة، حيث إن نماذج نظرية استجابة الفقرة (IRT) تنقسم إلى نوعين، هما: النماذج البارامترية والنماذج اللابارامترية، ولكل نوع من هذه الأنواع افتراضاته وإحصاءاته الخاصة به، حيث تتطلب النماذج البارامترية افتراضات أكثر يصعب تحقيقها مقارنة بنظيرتها اللابارامترية، مثل: وقوع البيانات على مقياس فئوي، وشكل دالة الفقرة، إضافة إلى ذلك إن نماذج نظرية استجابة الفقرة اللابارامترية تحتاج إلى حجم عينة وطول اختبار أقل من نظيرتها البارامترية، إلا أن المفاضلة بين النموذجين تكون في دقة التقدير لمعالم الفقرة ومعلمة القدرة التي يعطيها كلا النموذجين.

أهمية الدراسة

تنبثق أهمية هذه الدراسة من الحاجة لمعرفة دقة التقديرات المختلفة (معالم الفقرة ومعلمة القدرة) لطرق تقدير نظرية استجابة الفقرة البارامترية واللابارامترية، الأمر الذي يسمح بتحليل البيانات باستخدام النماذج البارامترية حين الإيفاء بافتراضاتها، وتوافر حجم العينة، وطول الاختبار المناسبين لتطبيق هذه النماذج، أما إذا انتهكت هذه الافتراضات ولم تتوافر حجوم عينات وأطوال اختبار مناسبة، فهل استخدام النماذج اللابارامترية يعطي تقديرات دقيقة، دون افتراض الفئوية لبيانات قد تكون رتبيه في الواقع أو شبه فئوية، ودون الوقوع في انتهاك الافتراضات التي تشوه النتائج (إضافة إلى الميزات الأخرى التي تمتاز بها النماذج اللابارامترية، مثل سهولة المطابقة، وشكل دالة الفقرة، وأنها تحتاج إلى حجم عينة وطول اختبار أقل).

ويمكن تحديد مشكلة الدراسة من خلال الأسئلة الرئيسة الآتية:

1. هل توجد فروق دالة إحصائيا عند مستوى الدلالة $\alpha=0.05$ بين متوسطات مؤشرات دقة تقدير معالم الفقرة المقدرة تعزى لنوع النموذج المستخدم (بارامتري، ولابارامتري) باختلاف متغيري الدراسة (حجم العينة وطول الاختبار)؟

2. هل توجد فروق دالة إحصائيا عند مستوى الدلالة $\alpha=0.05$ بين متوسطات مؤشرات دقة تقدير القدرة θ المقدرة تعزى لنوع النموذج المستخدم (بارامتري، لابارامتري) باختلاف متغيري الدراسة (حجم العينة، وطول الاختبار)؟

3. ما مدى التوافق في تقدير معالم الفقرة والقدرة باختلاف النموذج (بارامتري، ولابارامتري) وفقاً لمتغيري الدراسة (حجم العينة، وطول الاختبار)؟

التعريفات الإجرائية

- **نماذج نظرية استجابة الفقرة البارامترية (PIRT)**: هي النماذج التي تحدد شكل دالة استجابة الفقرة، وتتطلب افتراضات أكثر تشددا، بالإضافة إلى أنها لا تصلح إلا للبيانات الفئوية، التي تستخدم برنامج (BILOG-MG) الذي يستخدم طريقة الأرجحية العظمى الهامشية (MML) في التقدير.

- **نماذج نظرية استجابة الفقرة اللابارامترية (NIRT)**: هي النماذج التي لا تحدد شكل دالة استجابة الفقرة، التي تتطلب افتراضات أقل تشدداً، وتصلح للبيانات الفئوية والرتيبة، التي تستخدم برنامج (TESTGRAF) الذي يستخدم طريقة (KERNEL SMOOTHING) في التقدير.

- **حجم العينة**: تعني عدد المفحوصين الذين سيطبق عليهم الاختبار، وفي هذه الدراسة سيتم استخدام أربع عينات بأحجام مختلفة.

- **طول الاختبار**: يعني عدد الفقرات التي يتكون منها الاختبار، وفي هذه الدراسة سيتم استخدام ثلاثة اختبارات بأطوال مختلفة.

- **معالم الاختبار Test Parameters**: هي قيم إحصائية يتم تقديرها باستخدام معادلات رياضية، وتشمل كلا من معالم الأفراد (القدرة)، ومعالم الفقرات (الصعوبة، والتمييز، والتخمين).

- **معلمة القدرة Ability**: اللوغاريتم الطبيعي لدالة الأرجحية القصوى Maximum Likelihood Function لإجابة المفحوص اجابة صحيحة على الفقرات التي تعتبر نقطة صفر التدريج عند صعوبتها.

- **معلمة الصعوبة Threshold**: مستوى القدرة الذي يقابل احتمال 0.50، للإجابة على الفقرة إجابة صحيحة عندما يكون معامل التخمين مساويا للصفر.

- **معلمة التمييز Slope**: ميل منحنى خصائص الفقرة، الذي يقابل النقطة التي تكون فيها علامة القدرة تساوي صعوبة الفقرة.

- **معلمة التخمين Asymptote**: هي خط المقاربة الأدنى Lower Asymptote من منحنى خصائص الفقرة، ويمثل احتمالية إجابة المفحوصين ذوي القدرة المتدنية على الفقرة إجابة صحيحة.

- **دقة التقدير (Accuracy of Estimation)**: تعبير يشير إلى جودة التقدير التي يميزها الاحتمالية الكبيرة في أن التقدير قريب من القيمة الحقيقية، حيث يمكن الوصول إلى ذلك باختيار التقدير غير المتحيز (unbiased estimator)، الذي يتصف تباينه بأنه أقل تباين من أي تقدير آخر غير متحيز، وذلك باستخدام الجذر التربيعي لمعدل مربعات الأخطاء (RMSE).

محددات الدراسة

1- أجريت هذه الدراسة باستخدام طريقة المحاكاة (Simulation)، بيانات مولدة.

2- فيما يتعلق بطرق التقدير المستخدمة في هذه الدراسة، فقد تم استخدام طريقة التقدير (MML)؛ لتقدير معالم الفقرة ومعلمة القدرة للمنحى البارامتري، وطريقـة التقـدير (تنعيم النواة) KERNAL SMOOTHING، لتقدير معالم الفقرة، ومعلمـة القـدرة للمنحى اللابارامترية.

3- فيما يتعلق بتوزيعات معالم الفقرة والقدرة المولدة في الدراسة الحالية، فقد تم تبنـي التوزيع الطبيعي (0،1) لمعلمة القدرة ومعلمة الصعوبة، والتوزيع المنتظم لمعلمتي التمييز والتخمين.

4- تم إجراء هذه الدراسة بتكرار واحد لعملية تقدير معالم الفقرة والقدرة.

الفصل الثاني

الدراسات السابقة

الفصل الثاني

الدراسات السابقة

يتضمن هذا الجزء الدراسات السابقة ذات العلاقة بموضوع الدراسة، التي تم الاطلاع عليها. وقد تم عرضها وفقاً للترتيب الزمني بدءاً بالأقدم فالأحدث، ومن ثم وضع الباحث تعقيباً عليها يبرز ما تتميز به الدراسة الحالية عنها. وقام الباحث بتناول العديد من الدراسات التي أشارت إلى موضوع دقة القياس، والبحث عن أفضل الوسائل للوصول بالقياس النفسي إلى درجة من الدقة بما يمكن من اتخاذ قرارات صائبة بالاعتماد على نتائج عملية القياس.

وتمت الإشارة إلى أساليب تقدير المعالم في نماذج نظرية استجابة الفقرة البارامترية واللابارامترية ومميزات كل أسلوب، ومحكات الدقة في تقدير المعالم، والعوامل التي تؤثر في دقة تقدير معالم الفقرة، ومعلمة القدرة، خصوصاً عدد الفقرات وحجم عينة المفحوصين، وأهمية دقة التقدير في بناء الاختبارات واتخاذ القرارات. وفيما يأتي استعراض لأبرز تلك الدراسات:

قام كل من ريكيس ومارك (Reckase & Mark, 1978) بإجراء دراسة باستخدام أسلوب المحاكاة هدفت إلى مقارنة دقة تقدير معلمة القدرة، ومعالم الفقرة في نموذج راش والنموذج اللوجستي الثلاثي، حيث بينت النتائج أن النموذج اللوجستي الثلاثي قد طابق بيانات الاختبار بشكل أفضل من نموذج راش، وأن دقة تقدير معلمة القدرة لنموذج راش أقل من دقة تقدير معلمة القدرة للنموذج اللوجستي الثلاثي. كما بينت الدراسة أن النموذج اللوجستي الثلاثي يحتاج حجم عينة أكبر لمعايرة الفقرات من نموذج راش، وكانت أبرز نتائج هذه الدراسة أن

هناك ارتباطاً عالياً بين تقديرات القدرة وفق النموذجين لمعظم البيانات، وأن نموذج راش يفضل استخدامه في حالة العينات الصغيرة.

كما أجرى **كنغما وتنفرغرت (Kingma & Tenvergert, 1985)** دراسة هدفت للكشف عن تطور الحفظ لدى عينة مكونة من (108) طالبا وطالبة من طلبة الروضة والمدرسة الابتدائية ممن أتموا (13) وظيفة محفوظات من منشورات بياجيه. وقد استخدمت الدراسة نموذج التجانس الاطرادي المعروف باسم نموذج موكن اللابارامتري. وأشارت نتائج الدراسة إلى أن تطبيق تحليل التدريج لموكن كان مفيدا في تحديد رتب الطلبة على الاختبار.

كما أجرى **رامسي** (Ramsey, 1991) دراسة في الولايات المتحدة الأمريكية هدفت إلى الكشف عن مدى استخدام نماذج كيرنال اللابارامترية، في تقدير رتب القدرة، ولتحقيق هدف الدراسة تم تطبيق طريقة كيرنال على مجموعة من الأمثلة التي تضم متغيرات ثنائية الاستجابة. وقد بينت الدراسة أن طريقة كيرنال من أهم الطرق اللابارامترية، كما بينت الدراسة عدم وجود فقد للبيانات تأثر على دقة تقدير رتب القدرة في المتغيرات ثنائية الاستجابة.

وفي دراسة أعدها فيتزباترك وآن (Fitzpatrick & Ann, 2001) هدفت فحص أثر طول الاختبارات، التي تتكون من فقرتين وأربع فقرات وثماني فقرات واثنتي عشرة فقرة وعشرين فقرة على ثبات الاختبار، حيث خصص لكل فقرة درجتين وأربع درجات وست درجات، وقد تم تشكيل عينات حجما (200، و500، و1000) مفحوص بهدف فحص أثرها (طول الاختبار، وحجم العينة، وطريقة التصحيح) على دقة تقدير المعالم، حيث استخدم أسلوب المحاكاة لأغراض هذه الدراسة، وذلك بتوليد استجابات المفحوصين على اختبار في الرياضيات، وقد تم استخدام معيار الجذر التربيعي لمعدل مربعات الفروق (RMSD)؛ بهدف

تقييم دقة تقديرات المعالم. وتوصلت الدراسة إلى نتائج عدة، منها: أن دقة تقدير المعالم عندما تم استخدام عينة حجمها (200) مفحوص كانت أقل من دقة تقدير المعالم عندما تم استخدام عينة حجمها (500) مفحوص. وكانت دقة تقدير المعالم عندما تم استخدام عينة حجمها (200) مفحوص وعينة حجمها (500) مفحوص أقل من دقة تقديرات المعالم عندما تم استخدام عينة حجمها (1000) مفحوص؛ أي أن دقة التقدير تزداد بازدياد حجم العينة.

وأجرى كونينغ وسيجستما وهامرز (Koning & Sijtsma & Hamers, 2002) دراسة بهدف المقارنة بين نموذجين بارامتريين وآخرين لابارامتريين من نماذج استجابة الفقرة للتعرف إلى الفائدة المرجوة منها في تحليل بيانات الاختبار. وقد قام الباحثون بتطبيق اختبار للاستنتاج الاستقرائي على عينة مكونة من (478) طالباً وطالبة من طلبة الصف الثالث الأساسي. وقد تم تحليل البيانات باستخدام النماذج الآتية: النموذج البارامتري أحادي المعلمة (راش)، ونموذج فيرهيلست البارامتري (Verhelst model)، ونموذج موكن الاطرادي، ونموذج موكن المضاعف الاطرادي. وأظهرت النتائج أفضلية للجمع بين النوعين من النماذج البارامترية واللابارامترية، إذ قدمت النماذج اللابارامترية تدريجات رتبيه للفقرات والأفراد، كما قدمت النماذج البارامترية معلومات مفيدة حول خصائص الفقرات بالإضافة لفائدتها في بعض الجوانب التطبيقية كمعايرة درجات الاختبار، فالنماذج بنوعيها البارامترية واللابارامترية قدمت معلومات مختلفة باستخدام إحصائيات مختلفة، حيث فضلت الدراسة الجمع بينهما لتحسين نوعية الاختبار وجودته.

وأجرى وانج وهاريس وروسوس (Wang, Harris& Roussos, 2003) دراسة هدفت إلى التعرف إلى أثر تعددية أبعاد الاختبار على دقة تقدير معلمة القدرة، ومعلمة الفقرة في اختبارات القبول في كليات القانون في الولايات المتحدة الأمريكية (اختبار LSAT). وتم

في هذه الدراسة استخدام (11) قسما من اختبار (LSAT) في الدراسة الحالية. وأشارت نتائج الدراسة إلى أن تعدد الأبعاد في الاختبار قد أثر في تخفيض دقة تقدير معلمة القدرة، ومعالم الفقرة في الاختبار. كما أشارت النتائج إلى أن هذا الانخفاض في دقة التقدير كان أكثر وضوحا في مجموعات القدرة الأكثر تباينا.

وأجرى **ميجر وبانيك** (Meijer & Baneke,2004) دراسة هدفت إلى توضيح فوائد نظرية استجابة الفقرة اللابارامترية لبناء وتحليل المقاييس الشخصية والنفسية، حيث ناقش الباحثان قابلية تطبيق نماذج نظرية استجابة الفقرة اللابارامترية لبناء وتحليل المقاييس الشخصية والنفسية واختلاف هذه النماذج مع نماذج استجابة الفقرة البارامترية. ولغايات جمع البيانات تم استخدام وتحليل بيانات من الترجمة الهولندية الرسمية لمقياس (MMPI-2)، الذي يتألف من (33) فقرة تقيس مستويات مختلفة من الاكتئاب، وتضمنت العينة (439) فرداً بمتوسط عمر (32.5) عاماً، وكان (69%) منهم من الذكور. وأظهرت نتائج الدراسة أنه عبر استخدام النماذج اللابارامترية لنظرية استجابة الفقرة، يمكن الحصول على معلومات حول أداء الفقرات الأكثر صعوبة مقارنة بالنماذج البارامترية. كما أظهرت النتائج أن تلك النماذج هي نماذج مفيدة في استكشاف بنية البيانات. وقد أوصى الباحثان عند تحليل بيانات مقاييس الشخصية والعلاج النفسي استخدام نماذج نظرية استجابة الفقرة اللابارامترية.

وأجرت **لاي ودونبار وكولن** (Lei, Dunbar and Kolen, 2004) دراسة في أيوا بالولايات المتحدة الأمريكية هدفت إلى مقارنة النموذج البارامتري لاختبارات الاختيار من متعدد وطريقة كيرنال اللابارامترية لتقدير منحنى خصائص الفقرة (ICC)، ولتحقيق هدف الدراسة تم استخدام معيار تطبيقي تمثل في الكشف عن مدى استقرار تقديرات المنحنى في حالة تمثيلها بيانياً وبشكل عشوائي. وتم دراسة أثر زيادة معلم التبسيط (h : بارامتر التهذيب

(Smoothing Parameter)، وهو يعتمد بشكل أساسي على عدد المفحوصين، ويساوي (1.1N) في برنامج TESTGRAF) على النموذج اللابارامتري، وأثر صغر حجم العينة على النموذجين، وقد أظهرت النتائج وجود اختلافات جزئية بالنسبة للاستقرار ضمن النموذج الواحد (داخل النموذج)، وكان تزايد الاختلافات بسيطاً الذي يعزى للنموذج، وأدى كل من النموذجين إلى نفس النتائج في دقة تقدير منحنى خصائص الفقرة (ICC).

كما قام كل من **سيجتسما وايمونز وبوميستر ونكليشك وروردا** (**Sijtsma, Emons, Bouwmeester, Nyklicek & Roorda, 2007)** بدراسة تحليلية لنظرية استجابة الفقرة اللابارامترية لأبعاد مقياس جودة ورفاهية الحياة لمنظمة الصحة العالمية (WHOQOL-Bref)، حيث سعت الدراسة للكشف عن قدرة نموذج موكن الاطرادي اللابارمتري في تقويم وإنشاء تدريجات أحادية البعد مستلة من المقياس الأصلي متعدد الأبعاد، الذي تم تحليله باستخدام نموذج الاستجابة المتدرج. ولتحقيق أغراض الدراسة تم تطبيق المقياس على مجموعتين متساويتين من الرجال والنساء تزيد أعمارهم عن (30) سنة، وبعد ذلك تمت مناقشة العيوب والمزايا لكلا النموذجين البارامتري واللابارامتري، والبرمجيات المستخدمة في كلا النوعين، وبعد تحليل النتائج باستخدام النموذجين البارامتري واللابارامتري أنتج النموذج البارامتري تدريجات أحادية البعد لكل أبعاد المقياس الأصلي، كما أظهرت بعض الفقرات معاملات تدريج محدودة تبعاً لباقي الفقرات ضمن نفس التدريج، كما رفضت نتائج تحليل النموذج البارامتري بعض الفقرات، إلا أن النتائج النهائية أظهرت أن النموذج اللابارامتري للتجانس الاطرادي لموكن كان الأنسب والأكثر مطابقة لبيانات مقياس جودة ورفاهية الحياة.

وأجرى فو (Fu, 2010) دراسة بهدف التعرف إلى دقة تقدير معلمة القدرة ومعلمة صعوبة الفقرة باستخدام (5) نماذج من نظرية استجابة الفقرة البارامترية واللابارامترية، وكانت النماذج المستخدمة في إطار الدراسة متباينة من حيث مستوى التخمين، وأحجام العينة المستخدمة، وطول الاختبار. وتم في هذه الدراسة توليد مجموعة من الاستجابات بلغت (50) مجموعة من البيانات المولدة باستخدام ظروف اختبار مختلفة. وأشارت نتائج الدراسة إلى أن هناك تبايناً في دقة تقدير معالم الفقرة والمفحوصين حسب مستوى التخمين الموجود في الاختبار، وحجم العينة المستخدمة، وطول الاختبار. كما أشارت النتائج أيضا إلى أن دقة تقدير معلمة القدرة، ومعالم الفقرة تعتمد على معيار الدقة المستخدم في كل واحد من نماذج استجابة الفقرة.

وأجرى دي لا توري و يوان (De La Torre & Yuan, 2010) دراسة هدفت إلى التعرف إلى أثر حجم العينة على دقة تقدير معلمة القدرة ومعالم الفقرة في اختبارات مطورة حسب نماذج نظرية استجابة الفقرة (IRT). واستخدمت الدراسة نموذج (HO–IRT) في توليد مجموعة من البيانات ضمن ظروف اختبار مختلفة. حيث تم توليد البيانات باستخدام طريقة مونتي كارلو من أجل التعرف إلى العلاقات بين معلمة القدرة ومعالم الفقرة في الاختبار، وأثر حجم العينة في دقة تقدير معلمة القدرة ومعالم الفقرة. وأشارت النتائج إلى أن حجم العينة يؤثر في قدرة الاختبار في تقدير معلمة القدرة ومعالم الفقرة، وإلى وجود علاقة ارتباطية بين تقدير معلمة القدرة ومعالم الفقرة.

في حين أجرى سشمت وآخرون (Schmitt ,Sass, et al., 2010) دراسة هدفت إلى التعرف إلى دقة تقدير معلمة القدرة في الاختبارات التكيفية الحاسوبية. واستخدمت الدراسة منهجية مونتي كارلو، حيث تم توليد مجموعة من البيانات الخاصة بثلاثة اختبارات

تكيفيه لغوية. واستخدمت الدراسة ظروف اختبار مختلفة تباينت بين استخدام عدد فقرات مختلف (الفروق في طول الاختبار)، واستخدام سرعات إجابة مختلفة. وأشارت نتائج الدراسة إلى أن تحيز الفقرة مرتبط مع سرعة الإجابة على فقرات الاختبار، كما أشارت نتائج الدراسة إلى أن هناك علاقة بين دقة تقدير معلمة القدرة ومعالم الفقرة وبين طول الاختبار.

كما أجرى الشريفين (2012) دراسة هدفت إلى الكشف عن أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي في الفيزياء من نوع الاختيار من متعدد بأربعة بدائل تكون بصورته النهائية من (33) فقرة. وطبق الاختبار على عينة الدراسة المكونة من (1000) طالب وطالبة من طلبة الصف الثاني الثانوي العلمي، وحللت النتائج وفق النموذج الثلاثي المعلمة باستخدام البرمجية (Bilog-Mg). وبينت النتائج وجود فروق ذات دلالة إحصائية ($\alpha = 0.05$) في متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة وطريقة التقدير. كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية ($\alpha = 0.05$) بين متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لطريقة التقدير، كما بينت النتائج عدم وجود فروق دالة إحصائيا بين معاملات الثبات المقدرة وفق نظرية الاستجابة للفقرة عند أحجام العينة المختلفة (100، و500، و1000). وأشارت النتائج إلى أن دقة تقديرات معلمة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية، وعينة الأفراد ذوي القدرة المتدنية عند استخدام طريقة بييز التوقع (EAP)، في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة باستخدام طريقة الأرجحية العظمى (MLE) بغض النظر عن حجم العينة.

التعقيب على الدراسات السابقة

ومن خلال الدراسات التي تم عرضها سابقاً يمكن استخلاص بعض الاتجاهات في النتائج التي أمكن التوصل إليها فيما يتعلق بدقة تقدير المعالم في بعض نماذج نظرية الاستجابة للفقرة وذلك على النحو الآتي:

1. دقة تقدير المعالم في نماذج نظرية الاستجابة للفقرة تتأثر بطول الاختبار وحجم عينة المفحوصين.
2. إن دقة تقدير المعالم تتغير بشكل واضح تبعاً للطريقة المستخدمة في تقدير المعالم والبرنامج الحاسوبي المستخدم في عملية التقدير.
3. لم تقم أي دراسة -في حدود علم الباحث- بفحص أثر تشكيلات مختلفة لحجم عينة المفحوصين، وطول الاختبار وطريقة التقدير (بارامترية، ولابارامترية) على دقة تقدير معالم الفقرة والقدرة، وهذا يظهر أهمية إجراء الدراسة الحالية.
4. إن الدراسات السابقة لم تستخدم دالة معلومات الفقرة والاختبار كمؤشر من مؤشرات دقة التقدير، ولا مؤشر الكفاءة النسبية للمقارنة بين الاختبارات بيد أن هذه الدراسة استخدمت هذه المؤشرات بالإضافة إلى مؤشرات الدقة الأخرى، مثل: التحيز (BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE) للإجابة عن أسئلة الدراسة.
5. تتفق الدراسة الحالية مع الدراسات السابقة في سعيها لتقدير دقة معالم الفقرة والقدرة باستخدام نظرية استجابة الفقرة البارامترية واللابارامترية، في حين تختلف الدراسة الحالية مع الدراسات السابقة في أنها تختبر دقة التقدير باختلاف حجم العينة، وطول الاختبار، ونوع النموذج الرياضي المستخدم (بارامتري واللابارامتري).
6. تهتم الدراسة الحالية بطبيعة التأثيرات التي يحدثها حجم عينة المفحوصين، وعدد الفقرات، والنموذج الرياضي المستخدم في تقدير المعالم على دقة تقديرات معالم الفقرة ومعلم القدرة.

الفصل الثالث

الطريقة والإجراءات

الفصل الثالث

الطريقة والاجراءات

يتضمن هذا الفصل وصفاً لعملية توليد البيانات في ضوء قيم افتراضية للمعالم الحقيقية للقدرات والفقرات، والتحقق من افتراض البعدية، ويتناول وصفاً للمعالجات الإحصائية المستخدمة للإجابة عن أسئلة الدراسة.

توليد البينات

تم استخدام المنهج التجريبي في هذه الدراسة بتوليد بيانات باستخدام المحاكاة (لأن هذا الأسلوب يمكننا من ضبط المتغيرات) من الاستجابات الثنائية (1، 0) لعينات تحاكي عينات المجتمع الأصلي بطريقة المونتي كارلو (Monte Carlo) بحجوم عينات وأطوال اختبار مختلفة، وذلك باستخدام برنامج (WinGen v.3)، بحيث يتراوح أحجام العينات من (100 إلى 1000) مفحوص، وذلك وفقاً للتصنيف التالي (100، و250، و500، و1000) مفحوص تحت افتراض التوزيع الطبيعي للقدرة بواقع متوسط حسابي مقداره (0)، وانحراف معياري مقداره (1)، واختبارات يتراوح طولها من (20–60) فقرة، وذلك وفقاً للتصنيف التالي (20، و40، و60) فقرة تحت افتراض التوزيع الطبيعي لصعوبة الفقرات بواقع متوسط حسابي مقداره (0)، وانحراف معياري مقداره (1)، والتوزيع المنتظم لمعلمة التمييز بواقع القيمة الابتدائية (0.4) والقيمة النهائية (1.2)، والتوزيع المنتظم لمعلمة التخمين بواقع القيمة الابتدائية (0.2) والقيمة النهائية (0.3) على افتراض أن الاختبار المُولدة بياناته هو اختبار الاختيار من متعدد وله أربعة بدائل، تلائم النموذج ثلاثي المعلم، والجدول 1 يبين المتوسطات الحسابية والانحرافات المعيارية الخاصة بمعالم الفقرة الحقيقية لكلٍ من معلمة التمييز والصعوبة والتخمين لكل حالة من الحالات المشمولة بالدراسة الناتجة عن تفاعل متغيري الدراسة (طول الاختبار، وحجم العينة).

جدول 1 المتوسطات الحسابية والانحرافات المعيارية لمعالم الفقرة الحقيقية وفقاً لمتغيري الدراسة (طول الاختبار، حجم العينة)

حجم العينة	الإحصائي	عدد الفقرات: 20 فقرة a	20 فقرة B	20 فقرة c	40 فقرة a	40 فقرة b	40 فقرة C	60 فقرة a	60 فقرة B	60 فقرة c
100 فرد	القيمة الصغرى	0.49	-2.58	0.20	0.40	-2.26	0.20	0.41	-2.02	0.20
	المتوسط الحسابي	0.89	-0.02	0.24	0.76	0.25	0.25	0.83	-0.03	0.25
	الانحراف المعياري	0.24	0.92	0.03	0.23	1.01	0.03	0.23	0.88	0.03
	القيمة العظمى	1.18	1.53	0.29	1.19	1.79	0.30	1.19	2.39	0.30
250 فرد	القيمة الصغرى	0.49	-2.58	0.20	0.40	-2.26	0.20	0.41	-2.02	0.20
	المتوسط الحسابي	0.89	-0.02	0.24	0.76	0.25	0.25	0.83	-0.03	0.25
	الانحراف المعياري	0.24	0.92	0.03	0.23	1.01	0.03	0.23	0.88	0.03
	القيمة العظمى	1.18	1.53	0.29	1.19	1.79	0.30	1.19	2.39	0.30
500 فرد	القيمة الصغرى	0.49	-2.58	0.20	0.40	-2.26	0.20	0.41	-2.02	0.20
	المتوسط الحسابي	0.89	-0.02	0.24	0.76	0.25	0.25	0.83	-0.03	0.25
	الانحراف المعياري	0.24	0.92	0.03	0.23	1.01	0.03	0.23	0.88	0.03
	القيمة العظمى	1.18	1.53	0.29	1.19	1.79	0.30	1.19	2.39	0.30
1000 فرد	القيمة الصغرى	0.49	-2.58	0.20	0.40	-2.26	0.20	0.41	-2.02	0.20
	المتوسط الحسابي	0.89	-0.02	0.24	0.76	0.25	0.25	0.83	-0.03	0.25
	الانحراف المعياري	0.24	0.92	0.03	0.23	1.01	0.03	0.23	0.88	0.03
	القيمة العظمى	1.18	1.53	0.29	1.19	1.79	0.30	1.19	2.39	0.30

والجدول 2 يبين معلم القدرة الحقيقية لكل حالة من الحالات المشمولة بالدراسة الناتجة عن تفاعل متغيري (طول الاختبار، حجم العينة).

جدول 2 المتوسطات الحسابية والانحرافات المعيارية لمعلم القدرة الحقيقية وفقاً لمتغيري الدراسة (طول الاختبار، وحجم العينة)

حجم العينة	الإحصائي	طول الاختبار: 20 فقرة	40 فقرة	60 فقرة
100 فرد	القيمة الصغرى	-1.75	-1.75	-1.75
	المتوسط الحسابي	-0.02	-0.03	-0.03
	الانحراف المعياري	0.80	0.80	0.80
	القيمة العظمى	1.43	1.43	1.43
250 فرد	القيمة الصغرى	-2.11	-2.07	-2.11
	المتوسط الحسابي	-0.01	-0.01	0.00
	الانحراف المعياري	0.92	0.91	0.91
	القيمة العظمى	2.78	2.78	2.78
500 فرد	القيمة الصغرى	-2.86	-2.80	-2.73
	المتوسط الحسابي	0.04	0.02	0.02
	الانحراف المعياري	0.95	0.95	0.95
	القيمة العظمى	2.37	2.37	2.37
1000 فرد	القيمة الصغرى	-3.25	-3.25	-2.39
	المتوسط الحسابي	0.04	0.05	0.05
	الانحراف المعياري	1.00	0.98	0.97
	القيمة العظمى	3.71	3.71	3.06

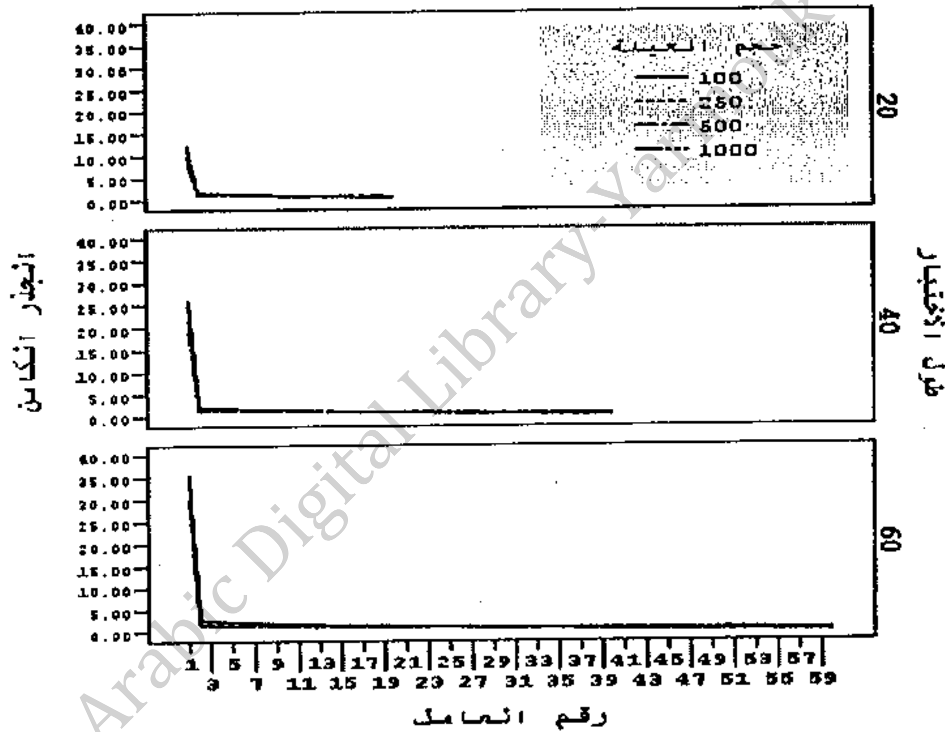
والجدول 3، يبين نتائج التحليل العاملي للبيانات المُوَلّدة عندما يكون طول الاختبار (20، و40، و60) وحجم العينة (100، و250، و500، و1000)، الذي تم استخدامه بهدف التحقق من افتراض أحادية البعد.

جدول 3: نتائج التحليل العاملي للبيانات المُوَلّدة وفقاً لمتغيري الدراسة (طول الاختبار، وحجم العينة)

حجم العينة	طول الاختبار	رقم العامل	الجذر الكامن	التباين المفسر %	التباين المفسر التراكمي	الجذر الكامن الأول / الجذر الكامن الثاني	(الجذر الكامن الأول - الجذر الكامن الثاني) / (الجذر الكامن الثاني - الجذر الكامن الثالث)
100	20	1	8.512	42.6	42.6	4.61	26.30
		2	1.848	9.2	51.8		
		3	1.595	8.0	**59.8**		
	40	1	24.746	61.9	61.9	11.23	259.75
		2	2.203	5.5	67.4		
		3	2.116	5.3	**72.7**		
	60	1	35.554	59.3	59.3	12.54	165.83
		2	2.836	4.7	64.0		
		3	2.639	4.4	**68.4**		
250	20	1	9.843	49.2	49.2	6.93	110.52
		2	1.420	7.1	56.3		
		3	1.344	6.7	**63.0**		
	40	1	19.015	47.5	47.5	10.78	214.88
		2	1.765	4.4	52.0		
		3	1.684	4.2	**56.2**		
	60	1	29.942	49.9	49.9	14.68	438.74
		2	2.040	3.4	53.3		
		3	1.976	3.3	**56.6**		
500	20	1	12.311	61.6	61.6	9.64	233.20
		2	1.277	6.4	67.9		
		3	1.230	6.1	**74.1**		
	40	1	26.111	65.3	65.3	17.74	621.76
		2	1.472	3.7	69.0		
		3	1.432	3.6	**72.5**		
	60	1	29.154	48.6	48.6	17.55	415.90
		2	1.661	2.8	51.4		
		3	1.595	2.7	**54.0**		
1000	20	1	10.985	54.9	54.9	9.49	190.19
		2	1.158	5.8	60.7		
		3	1.106	5.5	**66.2**		
	40	1	21.545	53.9	53.9	16.58	818.90
		2	1.299	3.2	57.1		
		3	1.274	3.2	**60.3**		
	60	1	33.951	56.6	56.6	24.73	1313.86
		2	1.373	2.3	58.9		
		3	1.348	2.2	**61.1**		

يلاحظ من الجدول 3، أن كافة قيم التباين المفسر عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (100، 250، 500، 1000) قد تخطت الـ 20% كمؤشر أوّل على أحادية البعد، كما وتخطت جميع قيم عملية قسمة الجذر الكامن الأول على الجذر الكامن الثاني قيمة

2 كمؤشر ثانٍ على تحقق أحادية البعد، وكذلك أظهرت نتائج عملية قسمة حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني قيماً ضخمة مما يشير إلى تحقق افتراض أحادية البعد كمؤشر ثالث، والشكل 1، يبين نتائج التحليل العاملي للبيانات المُولّدة وفقاً لمتغيري الدراسة (طول الاختبار، وحجم العينة) باستخدام الجذور الكامنة وعدد العوامل كمؤشر رابع على تحقق أحادية البعد.



شكل 1: رسم بياني يوضح نتائج التحليل العاملي للبيانات المولدة وفقاً لمتغيري الدراسة (طول الاختبار، وحجم العينة)

ومن ثم تم تقدير معالم الفقرات ومعلمة القدرة للمفحوصين باستخدام النموذج الثلاثي البارامتري باستخدام برنامج (Bilog-MG v.3)، حيث إن هذا البرنامج يستخدم أسلوب الأرجحية العظمى الهامشية (MML) في تقدير معالم الفقرة والقدرة، ومن ثم تم تقدير معالم الفقرات ومعلمة القدرة للمفحوصين باستخدام النموذج الثلاثي اللابارامتري باستخدام برنامج

(Testgraf version of July 2001)، حيث إن هذا البرنامج يستخدم أسلوب تنعيم النواة (Kernel Smoothing) في تقدير معالم الفقرة والقدرة، ومن ثم يتم المقارنة بين دقة تقدير النتائج للإجابة عن أسئلة الدراسة.

تحليل البيانات المولدة باستخدام برنامج Bilog-MG للنموذج الثلاثي البارامتري

تم تحليل البيانات المولدة باستخدام برنامج Bilog-MG للنموذج الثلاثي البارامتري، حيث تم تقدير معلمة القدرة θ للأفراد عندما يكون طول الاختبـار (20، 40، 60) وحجـم العينة (100، 250، 500، 1000).

وكذلك تم تقدير معالم (a, b, c) لفقرات الاختبار باستخدام النموذج البارامتري عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250، 500، 1000).

تحليل البيانات المولدة باستخدام برنامج Testgraf للنموذج الثلاثي اللابارامتري

تم تحليل البيانات المولدة باستخدام برنامج Testgraf للنموذج الثلاثـي اللابـارامتري، حيث تم تقدير معلمة القدرة θ للأفراد عندما يكون طول الاختبـار (20، 40، 60) وحجـم العينة (100، 250، 500، 1000).

وكذلك تم تقدير معالم (a, b, c) لفقرات الاختبار باستخدام النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250، 500، 1000).

المعالجات الإحصائية

تم استخدام المعالجات الإحصائية الآتية:

للإجابة عن سؤال الدراسة الأول؛ تم حساب المتوسطات الحسابية والانحرافات المعيارية لمؤشر دقة القياس ممثلةً بـ (BIAS) الخاصة بمعلم القدرة المقدرة باستخدام النموذجين (البارامتري، واللابارامتري) حيث تم حساب مؤشر الدقة BIAS باستخدام المعادلة $Bias(\theta^*) = \sum_{i=1}^{N}(\theta_i^* - \theta_i)/N$، متبوعة بإجراء تحليل التباين الثلاثي للقياسات المتكررة للكشف عن مواطن الدقة في كل حالات الدراسة، كما تم حساب المتوسطات الحسابية لمؤشر دقة القياس ممثلةً بـ (RMSE) الخاصة بمعلم القدرة المقدرة باستخدام النموذجين (البارامتري، واللابارامتري)، وتم حساب مؤشر الدقة RMSE (Root Mean Square Error) باستخدام المعادلة $RMSE(\theta^*) = \sqrt{\sum_{i=1}^{N}(\theta_i^* - \theta_i)^2/N}$، حيث θ_i^* القدرة المقدرة للمفحوص *i*، θ_i القدرة الحقيقية للمفحوص *i*، *N* عدد المفحوصين (Weiss, 2009).

كما تم حساب المتوسطات الحسابية والانحرافات المعيارية لدالة معلومات الاختبار المقدرة باستخدام النموذجين (البارامتري، واللابارامتري) في كل حالات الدراسة متبوعة بإنشاء رسوم بيانية توضحها، علاوة على حساب الكفاءة النسبية لدالتي معلومات الاختبار المقدرة وفقاً للنموذجين (البارامتري، واللابارامتري) إلى دالة معلومات الاختبار الحقيقية.

وللإجابة عن سؤال الدراسة الثاني؛ تم حساب المتوسطات الحسابية والانحرافات المعيارية لمؤشر دقة القياس ممثلةً بـ (BIAS) الخاصة بمعالم فقرات الاختبار (a, b, c) المقدرة باستخدام النموذجين (البارامتري، واللابارامتري) حيث تم حساب مؤشر الدقة BIAS باستخدام المعادلة $Bias(a,b,c^*) = \sum_{i=1}^{N}(a,b,c_i^* - a,b,c_i)/N$، متبوعة بإجراء تحليل التباين الثلاثي للقياسات المتكررة للكشف عن مواطن الدقة في كل حالات الدراسة، كما

تم حساب المتوسطات الحسابية لمؤشر دقة القياس ممثلةً بـــ (RMSE) الخاصة بمعالم فقرات الاختبار (a, b, c) المقدرة باستخدام النموذجين (البارامتري، واللابارامتري)، حيث تم حساب مؤشـــر الدقـــة (RMSE (Root Mean Square Error باســـتخدام المعادلـــة $RMSE(a,b,c^*) = \sqrt{\sum_{i=1}^{N}(a,b,c_i^* - a,b,c_i)^2/N}$، حيث a, b, c_i^* هـــي معـــالم الفقرات (a, b, c) المقدرة للمفحوص i، a, b, c_i هي معالم الفقرات الحقيقية للمفحوص i، N عدد المفحوصين.

وللإجابة عن سؤال الدراسة الثالث؛ تم حساب معاملات الارتباط باســتخدام معامــل ارتباط بيرسون كمؤشر على درجة الثقة لمعلمة القدرة ومعالم الفقرات المقدرة وفقاً للنموذجين (البارامتري، واللابارامتري) في كل حالات الدراسة مع معلمــة القــدرة ومعــالم الفقــرات الحقيقية.

الفصل الرابع
عرض النتائج ومناقشتها والتوصيات الخاصة بها

الفصل الرابع

عرض النتائج ومناقشتها والتوصيات الخاصة بها

يتناول هذا الفصل عرضاً لنتائج الدراسة مبوبة حسب أسئلتها:

أولاً. النتائج المتعلقة بالإجابة عن سؤال الدراسة الأول: "هل توجد فروق دالة إحصائيا عند مستوى الدلالة $\alpha=0.05$ بين متوسطات مؤشرات دقة تقدير معالم الفقرة المقدرة تعزى لنوع النموذج المستخدم (بارامتري، ولابارامتري) باختلاف متغيري الدراسة (حجم العينة، وطول الاختبار)؟"

وللإجابة عن هذا السؤال تم حساب المتوسطات الحسابية والانحرافات المعيارية لتقديرات معالم الفقرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 4.

جدول 4 المتوسطات الحسابية والانحرافات المعيارية لتقديرات معالم الفقرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

حجم العينة	الإحصائي	نوع النموذج	طول الاختبار: 20 فقرة a	20 فقرة b	20 فقرة C	40 فقرة A	40 فقرة B	40 فقرة C	60 فقرة a	60 فقرة b	60 فقرة C
100 فرد	المتوسط الحسابي	بارامتري	0.724	-0.262	0.209	0.585	-0.027	0.142	0.632	-0.005	0.245
		لابارامتري	0.414	2.705	0.452	0.782	-1199757.295	0.383	0.776	2.905	0.453
	الانحراف المعياري	بارامتري	0.19	1.42	0.02	0.18	1.41	0.01	0.19	1.10	0.02
		لابارامتري	0.73	5.70	0.23	0.90	7587948.93	0.29	0.95	9.25	0.27
250 فرد	المتوسط الحسابي	بارامتري	0.772	-0.693	0.241	0.697	0.272	0.221	0.821	0.125	0.274
		لابارامتري	0.732	1.852	0.441	0.481	28.497	0.320	0.526	1.317	0.407
	الانحراف المعياري	بارامتري	0.33	3.95	0.06	0.24	1.30	0.02	0.23	0.90	0.04
		لابارامتري	0.92	2.85	0.25	0.69	153.65	0.18	0.50	2.01	0.24
500 فرد	المتوسط الحسابي	بارامتري	0.976	0.121	0.284	0.894	0.578	0.325	0.781	-0.056	0.231
		لابارامتري	0.220	-758199.424	0.219	0.329	0.910	0.299	0.386	1.473	0.273
	الانحراف المعياري	بارامتري	0.25	0.99	0.04	0.30	1.09	0.06	0.24	0.95	0.04
		لابارامتري	0.14	3390773.62	0.16	0.27	2.04	0.21	0.45	5.57	0.23
1000 فرد	المتوسط الحسابي	بارامتري	1.118	0.304	0.328	0.921	0.561	0.322	1.054	0.262	0.330
		لابارامتري	0.507	0.898	0.444	0.427	0.574	0.306	0.485	0.475	0.342
	الانحراف المعياري	بارامتري	0.29	0.88	0.06	0.28	0.92	0.05	0.35	0.71	0.05
		لابارامتري	0.25	1.05	0.23	0.25	1.35	0.19	0.26	1.58	0.18

يلاحظ من الجدول 4، أن النتائج الخاصة به (المتوسطات الحسابية والانحرافات المعيارية) قد صُنِّفت في ضوء الإحصائي المستخدم، وذلك على النحو الآتي:

أ. **فيما يخص المتوسط الحسابي**: إن القيم المطلقة لكافة قيم المتوسط الحسابي لتقديرات معالم الفقرات (a, b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (a, b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (40، 60) وحجم العينة 100، كما يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 100، في حين أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معلمة الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لتقديرات معلمة الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 100.

كما يلاحظ من الجدول 4، أن القيم المطلقة لكافة قيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة 250. كما يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة 250.

وكذلك يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (a, c) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي

لتقديرات معالم الفقرات (a, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40) وحجم العينة 500. كما يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (b) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معلمة الفقرات (b) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40) وحجم العينة 500، في حين أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار 60 وحجم العينة 500، وأن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار 60 وحجم العينة 500.

وكذلك يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 60) وحجم العينة 1000. كما يلاحظ من الجدول 4، أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معلمة الفقرات (c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لتقديرات معلمة الفقرات (c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 60) وحجم العينة 1000، في حين أن القيم المطلقة لقيم المتوسط الحسابي لتقديرات معالم الفقرات (a, c) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (a, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار 40 وحجم العينة 1000، وأن القيم المطلقة لقيم المتوسط الحسابي لتقديرات

معالم الفقرات (b) حسب النموذج البارامتري قد كانت أصغر ظاهرياً من قيم المتوسط الحسابي لتقديرات معالم الفقرات (b) حسب النموذج اللابارامتري عندما يكون طول الاختبار 40 وحجم العينة 1000.

ب. **فيما يخص الانحراف المعياري:** يلاحظ من الجدول 4، أن كافة قيم الانحراف المعياري لتقديرات معالم الفقرات (a, b, c) حسب النموذج البارامتري قد كانت أصغر ظاهرياً من قيم الانحراف المعياري لتقديرات معالم الفقرات (a, b, c) حسب النموذج اللابارامتري في أغلب حالات الدراسة، باستثناء أن قيم الانحراف المعياري لتقديرات معلمة الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم الانحراف المعياري لتقديرات معلمة الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة 1000، وقيم الانحراف المعياري لتقديرات معلمة الفقرات (b) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم الانحراف المعياري لتقديرات معلمة الفقرات (b) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 250، وأن قيم الانحراف المعياري لتقديرات معالم الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم الانحراف المعياري لتقديرات معلمة الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40)، وحجم العينة 500.

وكذلك تم حساب المتوسطات الحسابية والانحرافات المعيارية الخاصة بمؤشر التحيز BIAS كمؤشر من مؤشرات دقة التقدير لمعالم الفقرات باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري)، وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 5.

جدول 5 المتوسطات الحسابية والانحرافات المعيارية لمؤشر التحيز لتقديرات معالم الفقرات باختلاف نوع النموذج المستخدم(بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة وطول الاختبار)

حجم العينة	الإحصائي	نوع النموذج	طول الاختبار: 20 فقرة (مؤشر التحيز لمعالم الفقرات:)			40 فقرة			60 فقرة		
			a	b	c	A	B	c	a	b	c
100 فرد	المتوسط الحسابي	بارامتري	-0.16	-0.24	-0.03	-0.17	-0.28	-0.11	-0.19	0.06	-0.01
		لابارامتري	-0.47	2.72	0.21	0.02	-1199757.55	0.13	-0.05	2.94	0.20
	الانحراف المعياري	بارامتري	0.24	0.61	0.03	0.25	0.71	0.03	0.23	0.47	0.03
		لابارامتري	0.69	5.55	0.23	0.92	7587949.12	0.29	1.02	9.20	0.27
250 فرد	المتوسط الحسابي	بارامتري	-0.11	-0.67	0.00	-0.06	0.02	-0.03	-0.01	0.16	0.02
		لابارامتري	-0.15	1.87	0.20	-0.28	28.24	0.07	-0.30	1.35	0.16
	الانحراف المعياري	بارامتري	0.30	3.28	0.06	0.21	0.48	0.03	0.21	0.35	0.04
		لابارامتري	1.03	3.01	0.24	0.71	153.48	0.18	0.54	1.94	0.24
500 فرد	المتوسط الحسابي	بارامتري	0.09	0.14	0.04	0.13	0.32	0.07	-0.05	-0.02	-0.019
		لابارامتري	-0.67	-758199.40	-0.02	-0.43	0.66	0.05	-0.44	1.51	0.02
	الانحراف المعياري	بارامتري	0.19	0.25	0.05	0.24	0.46	0.05	0.16	0.28	0.04
		لابارامتري	0.189	3390773.74	0.16	0.35	1.60	0.21	0.51	5.63	0.22
1000 فرد	المتوسط الحسابي	بارامتري	0.23	0.32	0.09	0.16	0.31	0.07	0.23	0.30	0.08
		لابارامتري	-0.38	0.92	0.20	-0.33	0.32	0.05	-0.34	0.51	0.09
	الانحراف المعياري	بارامتري	0.23	0.35	0.05	0.14	0.34	0.05	0.24	0.31	0.05
		لابارامتري	0.32	1.14	0.23	0.35	1.14	0.18	0.31	1.33	0.18

يلاحظ من الجدول 5، أن النتائج الخاصة به (المتوسطات الحسابية والانحرافات المعيارية لمؤشر التحيز) قد صُنِّفت في ضوء الإحصائي المستخدم، وذلك على النحو الآتي:

أ. **فيما يخص المتوسط الحسابي:** أن القيم المطلقة لكافة قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (a, b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (a, b, c) حسب النموذج اللابارامتري في أغلب الحالات باستثناء أن قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معلمة الفقرات (a) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معلمة الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار (40، 60) وحجم العينة 100، والقيم المطلقة لقيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (c) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40) وحجم العينة 500، والقيم المطلقة لقيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (c) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط الحسابي لمؤشر التحيز لتقديرات معالم الفقرات (c) حسب النموذج اللابارامتري عندما يكون طول الاختبار 40 وحجم العينة 1000)؛ مما يُعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معالم

الفقرات (a, b, c) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري في أغلب الحالات. وقد يعزى ذلك إلى أن طريقة التقدير اللابارامترية (KS) غالت في تقدير المعالم، وذلك أن هذه الطريقة لا تستطيع تقدير معالم الفقرات التي أجاب عليها كل المفحوصين أو لم يجب عليها أي مفحوص بعكس الطريقة البارامترية (MML)، التي تستطيع تقدير معالم الفقرة التي أجاب أو لم يجيب عليها كل المفحوصين، إضافة إلى ذلك قد يعزى السبب إلى أن معالم الفقرات المولدة (الحقيقية) أقرب لمستوى القياس الفئوي منه إلى الرتبي وذلك أن توزيع معلمة القدرة كان توزيعا طبيعيا (1, 0)، وهو احد افتراضات طريقة التقدير البارامارية MML .

ب. **فيما يخص الانحراف المعياري:** يلاحظ من الجدول 5، أن كافة قيم الانحراف المعياري لمؤشر التحيز لتقديرات معالم الفقرات (a, b, c) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم الانحراف المعياري لمؤشر التحيز لتقديرات معالم الفقرات (a, b, c) حسب النموذج اللابارامتري، باستثناء قيم الانحراف المعياري لمؤشر التحيز لتقديرات معلمة الفقرات (b) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم الانحراف المعياري لمؤشر التحيز لتقديزات معلمة الفقرات (b) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 250، وقيم الانحراف المعياري لمؤشر التحيز لتقديرات معلمة الفقرات (a) حسب النموذج البارامتري، قد كانت أكبر ظاهرياً من قيم الانحراف المعياري لمؤشر التحيز لتقديرات معلمة الفقرات (a) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 500، مما يُعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معالم الفقرات (a, b, c) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى ذلك إلى أن طريقة التقدير اللابارامترية (KS) غالت في تقدير المعالم، وذلك أن هذه الطريقة لا تستطيع تقدير معالم الفقرات التي أجاب عليها أو لم يجيب عليها كل الأفراد، ولذلك تقدر لها قيم متحيزة سواءً للأعلى أو للأسفل، بعكس الطريقة البارامترية (MML) التي تستطيع تقدير معالم الفقرة التي أجاب أو

لم يجيب عليها جميع الأفراد، إضافة إلى ذلك قد يعزى السبب إلى أن معالم الفقرات المولدة (الحقيقية) أقرب لمستوى القياس الفئوي منه إلى الرتبي، وذلك أن توزيع معلمة القدرة كان (1، 0) .

وفي ضوء ما تقدم، تم إجراء تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)، وذلك كما هو مبين في الجدول 6.

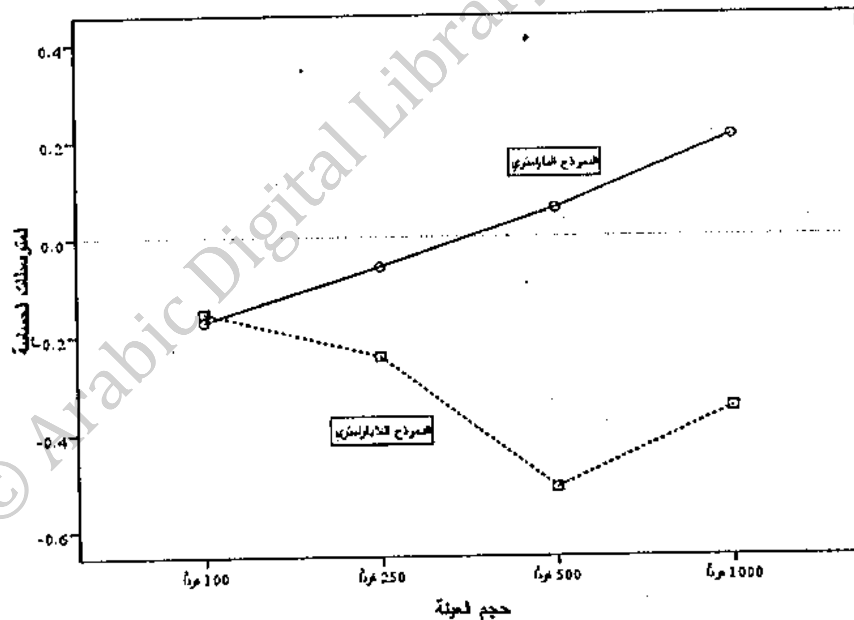
جدول 6 نتائج تحليل التباين الثلاثي للقياسات المتكررة لقيم مؤشر التحيز لتقديرات معلمة الفقرات (a) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)

الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
بين المجموعات	طول الاختبار	0.771	2	0.386	1.531	0.217
	حجم العينة	2.378	3	0.793	**3.147**	0.025
	طول الاختبار×حجم العينة	1.715	6	0.286	1.135	0.341
	الخطأ	117.379	466	0.252		
	الكلي	122.244	477			
داخل المجموعات	النموذج	20.555	1	20.555	**99.521**	0.000
	طول الاختبار× النموذج	0.830	2	0.415	2.010	0.135
	حجم العينة× النموذج	12.371	3	4.124	**19.965**	0.000
	طول الاختبار×حجم العينة× النموذج	2.837	6	0.473	**2.289**	0.035
	الخطأ	96.249	466	0.207		
	الكلي	132.842	478			
الكلي		255.086	955			

ويلاحظ من الجدول 6 وجود فرق دال إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطين الحسابيين لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة تعزى لنوع النموذج المستخدم (بارامتري، لابارامتري)؛ لصالح النموذج البارامتري مقارنة بالنموذج اللابارامتري. وقد يعزى السبب إلى الأساس الرياضي لكل من الطريقتين (البارامترية واللابارامترية) حيث إن طريقة التقدير اللابارامترية (KS) تعتمد في تقديرها لمنحنى خصائص الفقرة على الرتب في تقدير منحنى خصائص الفقرة، ومن ثم تقدر معلمة

التمييز(a)، وهذا التحويل من مستوى القياس الفئوي إلى مستوى القياس الرتبي يؤدي إلى خسارة في المعلومات مما يؤدي إلى انخفاض دقة التقدير، وعلى العكس في طريقة التقدير البارامترية (MML) حيث إنها تقدر معلمة التمييز مباشرة.

ويتضح من الجدول 6 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة)، والشكل 2 يوضح التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة.

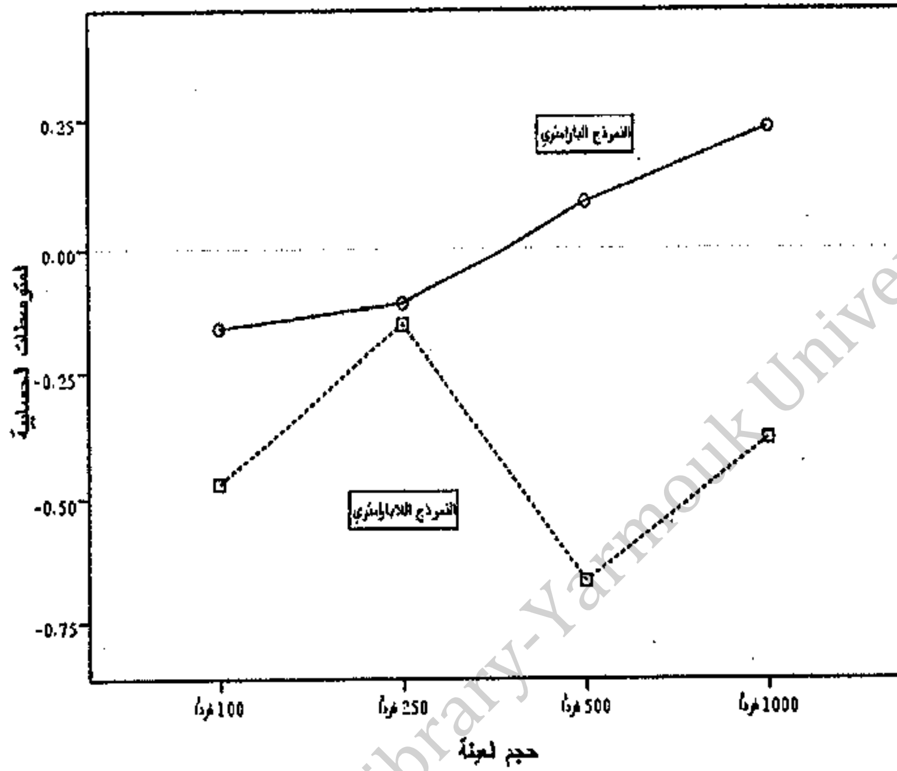


شكل 2: التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة.

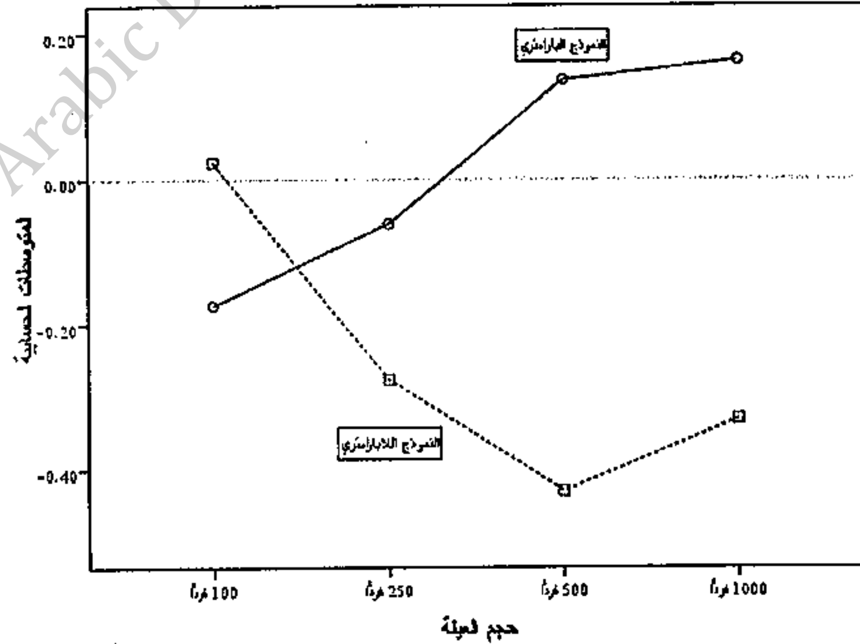
يلاحظ من الشكل 2 أنه تفاعل لا رتبي لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة)، لصالح النموذج البارامتري، عندما يكون حجم العينة (250، 500، 1000) فرداً؛

مما يُعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري. وقد يعزى ذلك إلى أن طريقة التقدير اللابارامترية (KS) غالت في تقدير معلمة التمييز(a)، وذلك أن هذه الطريقة تقوم بتحويل الدرجات الخام إلى رتب، ومن ثم تقوم بتقدير منحنى خصائص الفقرة، ثم تقدر المعالم، مما يؤدي إلى فقد في المعلومات، وبالتالي تقل دقة التقدير له مقارنة بالنموذج البارامتري الذي يستخدم طريقة (MML) في تقدير معالم الفقرة والقدرة. إضافة إلى ذلك قد يعزى السبب إلى أن توزيع معلمة القدرة الحقيقية (المولدة) أقرب لمستوى القياس الفئوي منه إلى الرتبي، حيث إنها تتوزع توزيعا طبيعيا (1 ، 0)، وهو افتراض من افتراضات طريقة التقدير البارامترية MML، وبالتالي فإنها تكون مناسبة للنموذج البارامتري أكثر منه للنموذج اللابارامتري، وهي لصالح النموذج اللابارامتري عندما يكون حجم العينة (100) فرداً؛ مما يُعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، وهذه النتيجة تتوافق مع ما ادعاه رامسي (Ramsay,1991) أنه يمكن تقدير معالم الفقرات باستخدام عدد قليل من الفقرات والمفحوصين 20 فقرة و100 فرد، وان طريقة التقدير اللابارامترية KS مناسبة لتقدير معلمة القدرة عند حجوم العينات الصغيرة .

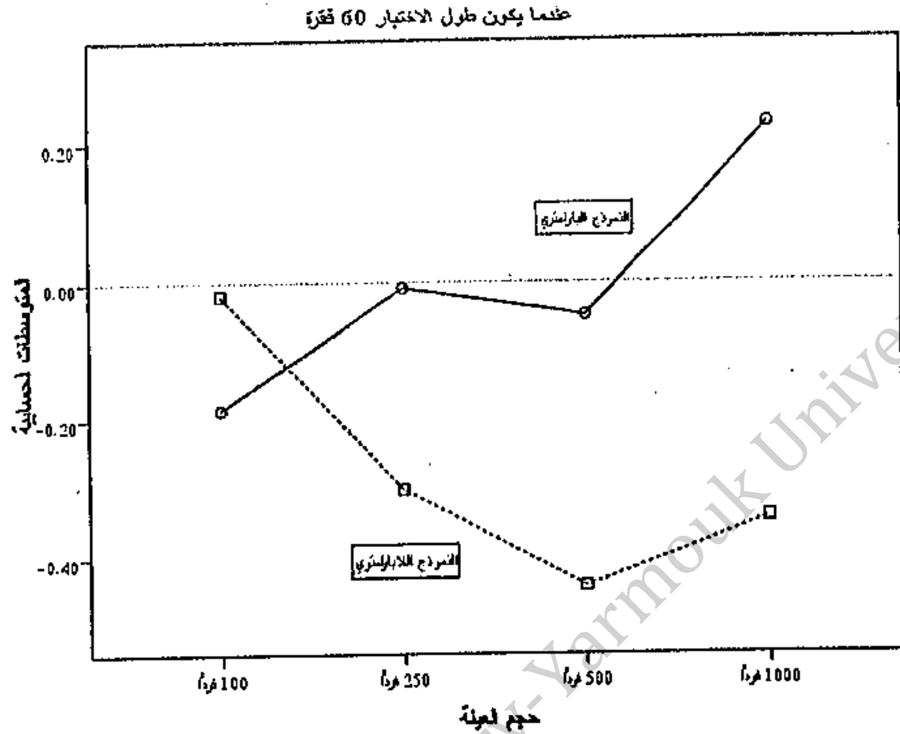
كما يتضح من الجدول 6 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة)، والأشكال (3، 4، 5) توضح التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة) على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة.



شكل 3: التفاعل لنوع النموذج المستخدم (بارامتري، لابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 20 فقرة على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة



شكل 4: التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 40 فقرة على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة



شكل 5: التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 60 فقرة على قيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة

يلاحظ من الشكل 3 أنه تفاعل رتبي لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (a) المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) عندما كانت حجوم العينات (100، 250، 500، 1000) فرداً، وعندما يكون طول الاختبار 20 فقرة، لصالح النموذج البارامتري، مما يُعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري. وقد يعزى ذلك إلى أن طريقة التقدير اللابارامترية (KS) غالت في تقدير معلمة التمييز(a)، وذلك أن هذه الطريقة تقوم بتحويل الدرجات الخام إلى رتب، ومن ثم تقوم بتقدير منحنى خصائص الفقرة، ثم تقدر المعالم، مما يؤدي إلى فقد في المعلومات، وبالتالي تقل دقة التقدير له مقارنة بالنموذج البارامتري الذي يستخدم طريقة (MML) في تقدير معالم الفقرة والقدرة. إضافة إلى ذلك قد يعزى السبب إلى أن توزيع معلمة القدرة الحقيقية(المولدة) أقرب لمستوى القياس الفئوي منه إلى الرتبي حيث إنها تتوزع توزيعا طبيعيا (1 , 0)، وبالتالي فإنها تكون مناسبة للنموذج

البارامتري أكثر منه للنموذج اللابارامتري.

كما يلاحظ من الشكلين (4، 5) أنهما تفاعلان لارتبيان لقيم مؤشـــر التحيـــز BIAS لتقديرات معلمة الفقرات (a) المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) عندما تكون حجوم العينات (250، 500، 1000) فرداً، وعندما يكون طول الاختبـــار (40، 60) فقرة، لصالح النموذج البارامتري، مما يُعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج اللابـــارامتري. وقـــد يعزى السبب إلى أن حجم العينة وطول الاختبار كبيران ومناسبان للنموذج البارامتري. وهذه النتيجة تتوافق مع ما بينه هلن، وليـــساك، ودراســـجو (Hulin, lissak, and Drasgow, 1982) المشار إليهم في (Crocker and Algena, 1986)، أن الدقة في تقدير معالم الفقرة الثلاثة في نماذج السمات الكامنة تزداد بزيادة حجم العينة من (200) إلـــى (500) ثـــم إلـــى (1000) على الترتيب، إلا أنه لم يتبين أي تحسن في دقة التقدير عند زيادتها إلى (2000).

ولصالح النموذج اللابارامتري عندما يكون حجم العينة (100) فرداً، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، وهذه النتيجة تتوافق مع ما ادعاه رامسي (Ramsay,1991) من أنه يمكن تقدير معالم الفقرات باستخدام عدد قليل من الفقرات والمفحوصين 20 فقرة و100 فرد.

وفي ضوء ما تقدم، تم إجراء تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (b) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)، وذلك كما هو مبين في الجدول 7.

جدول 7 نتائج تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز لتقديرات معلمة الفقرات (b) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)

الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
بين المجموعات	طول الاختبار	4461624783675.400	2	2230812391837.700	0.844	0.431
	حجم العينة	5742335895015.290	3	1914111965005.100	0.724	0.538
	طول الاختبار×حجم العينة	19429912163911.200	6	3238318693985.200	1.225	0.292
	الخطأ	1231975654788660.000	466	2643724581091.550		
	الكلي	1261609527631270.000	477			
داخل المجموعات	النموذج	5219232303549.980	1	5219232303549.980	1.974	0.161
	طول الاختبار× النموذج	4461620859522.930	2	2230810429761.470	0.844	0.431
	حجم العينة× النموذج	5742328838223.330	3	1914108946074.440	0.724	0.538
	طول الاختبار×حجم العينة× النموذج	19429899692930.900	6	3238316615488.480	1.225	0.292
	الخطأ	1231975831277730.000	466	2643724959823.460		
	الكلي	1266828910971960.000	478			
الكلي		2528438438603220.000	955			

يتضح من الجدول 7 عدم وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (b) المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة)، مما يعني عدم وجود فروق ذات دلالة احصائية تعزى لأثر طريقة التقدير (بارامتري، واللابارامتري)، باختلاف متغيري (حجم العينة، وطول الاختبار).

وفي ضوء ما تقدم، تم إجراء تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (c) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)، وذلك كما هو مبين في الجدول 8.

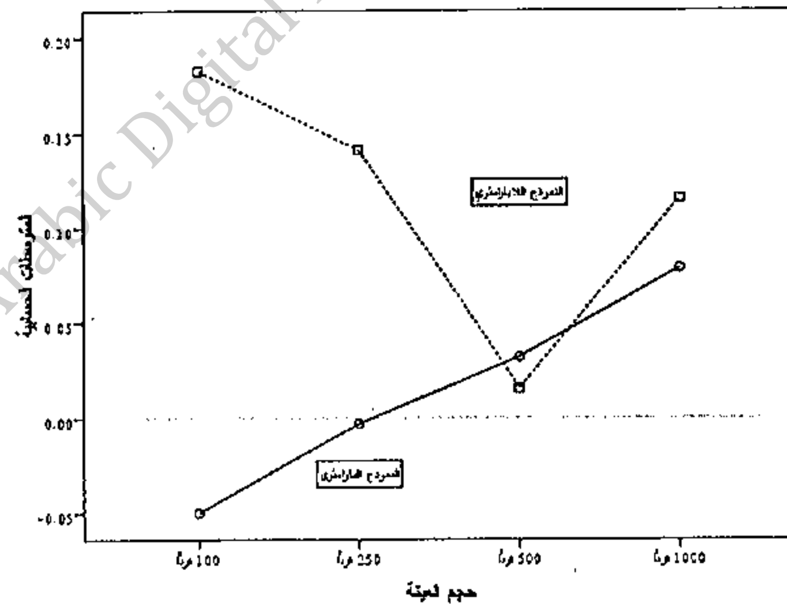
جدول 8 نتائج تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز لتقديرات معلمة الفقرات (c) المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)

الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
بين المجموعات	طول الاختبار	0.311	2	0.155	**5.055**	0.007
	حجم العينة	0.543	3	0.181	**5.889**	0.001
	طول الاختبار×حجم العينة	0.754	6	0.126	**4.090**	0.001
	الخطأ	14.315	466	0.031		
	الكلي	15.922	477			
داخل المجموعات	النموذج	1.939	1	1.939	**88.526**	0.000
	طول الاختبار× النموذج	0.070	2	0.035	1.604	0.202
	حجم العينة× النموذج	1.807	3	0.602	**27.500**	0.000
	طول الاختبار×حجم العينة× النموذج	0.242	6	0.040	1.839	0.090
	الخطأ	10.209	466	0.022		
	الكلي	14.268	478			
الكلي		30.190	955			

يتضح من الجدول 8 وجود فرق دال إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطين الحسابيين لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (c) المقدرة تعزى

لمتغير نوع النموذج المستخدم (بارامتري، ولابارامتري)، لصالح النموذج البارامتري مقارنة بالنموذج اللابارامتري. وقد يعزى ذلك إلى أن طريقة التقدير اللامعلمية (KS) غالت في تقدير معلمة التخمين(c)، وذلك أن هذه الطريقة تعتمد الترتيب في تقدير منحنى خصائص الفقرة، وأنها تعطي الرتب المتساوية رتب عشوائية؛ أي أنه لا يوجد تكرار لنفس الرتبة، ومن خلال منحنى خصائص الفقرة تقدر معلمة التخمين، مما يؤدي إلى فقد في المعلومات، وبالتالي تقل دقة التقدير له مقارنة بالنموذج البارامتري الذي يستخدم طريقة (MML) في تقدير معالم الفقرة والقدرة.

كما يتضح من الجدول 8 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (c) المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة)؛ والشكل 6 يوضح التفاعل لمتغيري النموذج المستخدم وحجم العينة على قيم مؤشر التحيز لتقديرات معلمة الفقرات (c) المقدرة.



شكل 6: التفاعل لمتغيري النموذج المستخدم وحجم العينة على قيم مؤشر التحيز لتقديرات معلمة الفقرات (c) المقدرة.

يلاحظ من الشكل 6 أنه تفاعل لارتبي لقيم مؤشر التحيز BIAS لتقديرات معلمة الفقرات (c) المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) عندما كانت حجوم

العينات (100، 250، 1000) فرداً لصالح النموذج البارامتري، مما يعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (c) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وعندما كانت حجوم العينات 500 فرداً لصالح النموذج اللابارامتري، مما يُعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة الفقرات (c) مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، وقد يعزى ذلك إلى أن طريقة التقدير اللامعلمية (KS) غالت في تقدير معلمة التخمين(c)، وذلك أن هذه الطريقة تعتمد الترتيب في تقدير منحنى خصائص الفقرة، وأنها تعطي الرتب المتساوية رتب عشوائية؛ أي أنه لا يوجد تكرار لنفس الرتبة، ومن خلال منحنى خصائص الفقرة تقدر معلمة التمييز، مما يؤدي إلى فقد في المعلومات، وبالتالي تقل دقة التقدير له مقارنة بالنموذج البارامتري الذي يستخدم طريقة (MML) في تقدير معالم الفقرة، إضافة إلى ذلك قد يعزى السبب إلى أن توزيع معلمة القدرة الحقيقية (المولدة) أقرب لمستوى القياس الفئوي منه إلى الرتبي حيث إنها تتوزع توزيعا طبيعيا (1 , 0)، وبالتالي فإنها تكون مناسبة للنموذج البارامتري أكثر منه للنموذج اللابارامتري.

كذلك في ضوء ما تقدم، فقد تم حساب قيم مؤشر RMSE كمؤشر من مؤشرات دقة التقدير لمعالم الفقرات باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 9.

جدول 9 قيم مؤشر RMSE لتقديرات معالم الفقرات باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

طول الاختبار	حجم العينة	RMSE لمعلمة التمييز وفقاً للنظرية استجابة الفقرة:		RMSE لمعلمة الصعوبة وفقاً للنظرية استجابة الفقرة:		RMSE لمعلمة التخمين وفقاً لنظرية استجابة الفقرة:	
		البارامترية	اللابارامترية	البارامترية	اللابارامترية	البارامترية	اللابارامترية
20 فقرة	100 فرد	0.287	0.825	0.643	6.057	0.045	0.308
	250 فرد	0.314	1.016	3.270	3.484	0.061	0.309
	500 فرد	0.208	0.692	0.281	3390773.597	0.060	0.161
	100 فرد	0.323	0.469	0.473	1.441	0.100	0.303
40 فقرة	100 فرد	0.300	0.912	0.756	7587948.714	0.114	0.313
	250 فرد	0.213	0.752	0.475	154.160	0.045	0.186
	500 فرد	0.276	0.549	0.556	1.714	0.091	0.208
	100 فرد	0.215	0.478	0.456	1.172	0.085	0.188
60 فقرة	100 فرد	0.291	1.008	0.467	9.587	0.033	0.337
	250 فرد	0.212	0.619	0.379	2.347	0.050	0.285
	500 فرد	0.167	0.670	0.277	5.783	0.045	0.223
	100 فرد	0.329	0.461	0.424	1.412	0.095	0.204

يلاحظ من الجدول 9، أن كافة قيم مؤشر دقة التقدير RMSE لمعالم فقرات الاختبار ذي البيانات المولّدة (a, b, c) حسب النموذج البارامتري، قد كانت **أصغر** ظاهرياً من قيم مؤشر دقة التقدير RMSE لمعالم فقرات الاختبار ذي البيانات المولّدة (a, b, c) حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معالم الفقرات (a, b, c) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى ذلك إلى الأساس الرياضي لكل من الطريقتين (البارامترية، واللابارامترية) حيث إن طريقة التقدير اللابارامترية (KS) تعتمد في تقديرها على الرتب، حيث إنها تحول الدرجات الخام إلى رتب، ثم تقدر منحنى خصائص الفقرة، ومن ثم تقدر قيم المعالم، وهذا يؤدي إلى خسارة في المعلومات، وبالتالي إلى انخفاض دقة التقدير، وعلى العكس في طريقة التقدير البارامترية (MML) حيث إنها تقدر معالم الفقرات (a, b, c) مباشرة، كما يتم في هذه الطريقة التدوير المتعاقب (Iteration) للوصول إلى أدق قيمة، والتي تعطي قيمة عظمى لاقتران الأرجحية العظمى، إضافة إلى أن توزيع المعالم الحقيقية (المولدة) توزيع طبيعي (1 ، 0) لمعلمة القدرة، وهو أنسب للنموذج البارامتري منه إلى النموذج اللابارامتري، إلا أن قيمة مؤشر دقة التقدير

RMSE لمعلمة فقرات الاختبار ذي البيانات المولّدة (b) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم مؤشر دقة التقدير RMSE لمعلمة فقرات الاختبار ذي البيانات المولّدة (b) حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة 100، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة الفقرات (b) مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، وقد يعزى ذلك إلى أن النموذج اللابارامتري لا يتأثر بشكل توزيع المعالم الحقيقية (المولدة) ولا يتأثر بصغر حجم العينة وطول الاختبار، وذلك لأنه يعتمد على الرتب في تقدير منحنى خصائص الفقرة، وبالتالي فهو أكثر دقة من النموذج البارامتري في حجوم العينات وأطوال الاختبارات الصغيرة، وهذا يتوافق مع الادب النظري في أن النموذج اللابارامتري يحتاج إلى حجوم عينات وأطوال اختبار أقل من النموذج البارامتري.

ثانياً. النتائج المتعلقة بسؤال الدراسة الثاني: "هل توجد فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين متوسطات مؤشرات دقة تقدير القدرة θ المقدرة تعزى لنوع النموذج المستخدم (بارامتري،واللابارامتري) باختلاف متغيري الدراسة (حجم العينة، وطول الاختبار)؟"

للإجابة عن هذا السؤال؛ تم حساب المتوسطات الحسابية والانحرافات المعيارية لقيم القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، والجدول 10 يبين هذه النتائج.

جدول 10 المتوسطات الحسابية والإنحرافات المعيارية لقيم القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

طول الاختبار	الإحصائي	نوع النموذج	حجم العينة 100 فرد	250 فرد	500 فرد	1000 فرد
20 فقرة	المتوسط الحسابي	البارامترية	0.012	-0.037	-0.023	-0.036
		اللابارامترية	0.584	0.311	0.214	-0.034
	الانحراف المعياري	البارامترية	1.41	1.50	1.41	1.40
		اللابارامترية	2.28	1.88	1.81	1.79
40 فقرة	المتوسط الحسابي	البارامترية	0.023	-0.007	-0.054	-0.050
		اللابارامترية	-0.096	0.022	0.040	-0.023
	الانحراف المعياري	البارامترية	1.25	1.25	1.31	1.31
		اللابارامترية	1.70	1.48	1.49	1.47
60 فقرة	المتوسط الحسابي	البارامترية	0.009	-0.005	-0.002	-0.031
		اللابارامترية	0.124	0.076	0.033	0.009
	الانحراف المعياري	البارامترية	1.17	1.16	1.16	1.20
		اللابارامترية	1.51	1.35	1.34	1.30

يلاحظ من الجدول 10، أن القيم المطلقة لقيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج اللابارامتري في أغلب الحالات، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء المتوسط الحسابي لمعلمة القدرة المقدرة وفقاً للنموذج البارامتري أقرب ما يمكن من المتوسط الحسابي المفترض في معرض توليد معلمة القدرة باستخدام برنامج WinGen، باستثناء القيم المطلقة لقيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار (40) وحجم العينة (500، 1000)، والقيم المطلقة لقيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسطات الحسابية لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار (60) وحجم العينة 1000، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج البارامتري، حيث جاء المتوسط الحسابي لمعلمة القدرة

المقدرة وفقاً للنموذج اللابارامتري أقرب ما يمكن من المتوسط الحسابي المفترض في معرض توليد معلمة القدرة باستخدام برنامج WinGen.

كما تم حساب المتوسطات الحسابية والانحرافات المعيارية لمؤشر التحيز BIAS كمؤشر من مؤشرات دقة التقدير لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 11.

جدول 11 المتوسطات الحسابية والإنحرافات المعيارية لمؤشر التحيز لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، لابارامتري) وباختلاف متغيري (حجم العينة وطول الاختبار)

طول الاختبار	الإحصائي	التحيز في القدرة باختلاف نوع النموذج:	حجم العينة 100 فرد	250 فرد	500 فرد	1000 فرد
20 فقرة	المتوسط الحسابي للتحيز	بارامتري	0.043	-0.018	-0.013	-0.059
		لابارامتري	0.615	0.329	0.224	-0.057
	الانحراف المعياري للتحيز	بارامتري	0.91	0.98	0.95	0.88
		لابارامتري	1.79	1.42	1.33	1.25
40 فقرة	المتوسط الحسابي للتحيز	بارامتري	0.054	0.011	-0.044	-0.072
		لابارامتري	-0.065	0.040	0.050	-0.046
	لانحراف المعياري للتحيز	بارامتري	0.81	0.77	0.79	0.75
		لابارامتري	1.21	1.02	0.97	0.92
60 فقرة	المتوسط الحسابي للتحيز	بارامتري	0.040	0.013	0.008	-0.054
		لابارامتري	0.155	0.094	0.042	-0.013
	الانحراف المعياري للتحيز	بارامتري	0.62	0.56	0.54	0.56
		لابارامتري	0.98	0.73	0.70	0.63

يلاحظ من الجدول 11، أن القيم المطلقة لقيم المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (100، 250، 500)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج البارامتري أقرب ما يمكن من الصفر. في حين يلاحظ من الجدول 11، أن القيم المطلقة لقيم المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المتوسط

الحسابي لمؤشر التحيز لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (1000)، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج البارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج اللابارامتري أقرب ما يمكن من الصفر.

كما يلاحظ من الجدول 11، أن كافة قيم الانحراف المعياري لمؤشر التحيز لتقديرات القدرة حسب النموذج البارامتري كانت **أصغر** ظاهرياً من قيم الانحراف المعياري لمؤشر التحيز لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء الانحراف المعياري لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج البارامتري أقل ما يمكن.

وفي ضوء ما تقدم، تم إجراء تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)، وذلك كما هو مبين في الجدول 12.

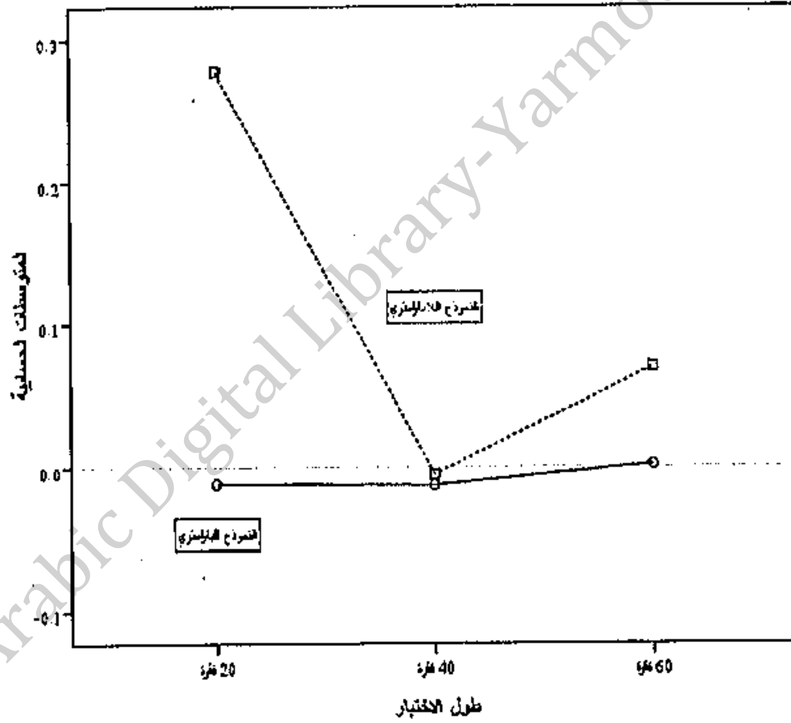
جدول 12 نتائج تحليل التباين ثلاثي التفاعل للقياسات المتكررة لقيم مؤشر التحيز **BIAS** لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (طول الاختبار، وحجم العينة)

الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	قيمة ف المحسوبة	الدلالة الإحصائية
بين الأفراد	طول الاختبار	19.822	2	9.911	**7.132**	0.001
	حجم العينة	41.831	3	13.944	**10.034**	0.000
	طول الاختبار×حجم العينة	15.471	6	2.579	1.855	0.085
	الخطأ	7696.219	5538	1.390		
	الكلي	7773.343	5549			
داخل الأفراد	النموذج	20.872	1	20.872	**85.616**	0.000
	طول الاختبار× النموذج	20.755	2	10.377	**42.567**	0.000
	حجم العينة× النموذج	9.942	3	3.314	**13.593**	0.000
	طول الاختبار×حجم العينة× النموذج	19.211	6	3.202	**13.133**	0.000
	الخطأ	1350.104	5538	0.244		
	الكلي	1420.883	5550			
الكلي		9194.226	11099			

يتضح من الجدول 12 وجود فرق دال إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطين الحسابيين لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة تعزى لمتغير نوع النموذج المستخدم (بارامتري، ولابارامتري)، لصالح النموذج البارامتري مقارنة بالنموذج اللابارامتري، مما يعني أن النموذج البارامتري أكثر دقة من النموذج اللابارامتري، وقد يعزى ذلك إلى الأساس الرياضي لكل من الطريقتين (البارامترية، واللابارامترية)، حيث إن طريقة التقدير اللابارامترية (KS) تعتمد في تقديرها على الرتب حيث إنها تحول الدرجات الخام إلى رتب، ثم تقدر منحنى خصائص الفقرة، ومن ثم تقدر قيمة معلمة القدرة (θ)، وهذا يؤدي إلى خسارة في المعلومات، وبالتالي إلى انخفاض دقة التقدير وعلى العكس في طريقة التقدير البارامترية (MML)، حيث إنها تقدر معالمة القدرة (θ) مباشرة، ومن ثم تقدر منحنى خصائص الفقرة اعتمادا على معالم الفقرة، كما يتم في هذه الطريقة القيام بالتدوير المتعاقب Iteration لعملية التقدير قد تصل إلى 100 تكرار أو أكثرمن اجل الحصول على تقديرات ثابتة لمعلمة القدرة (θ)، التي تعطي قيمة عظمى لاقتران الأرجحية العظمى، إضافة إلى أن توزيع المعالم الحقيقية (المولدة) توزيع طبيعي (1 , 0) لمعلمة الصعوبة وتوزيع منتظم

لمعلمتي التمييز والتخمين، وهو أنسب للنموذج البارامتري منه إلى النموذج اللابارامتري.

كما يتضح من الجدول 12 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (طول الاختبار)، والشكل 7 يوضح التفاعل الثنائي لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (طول الاختبار) على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة.

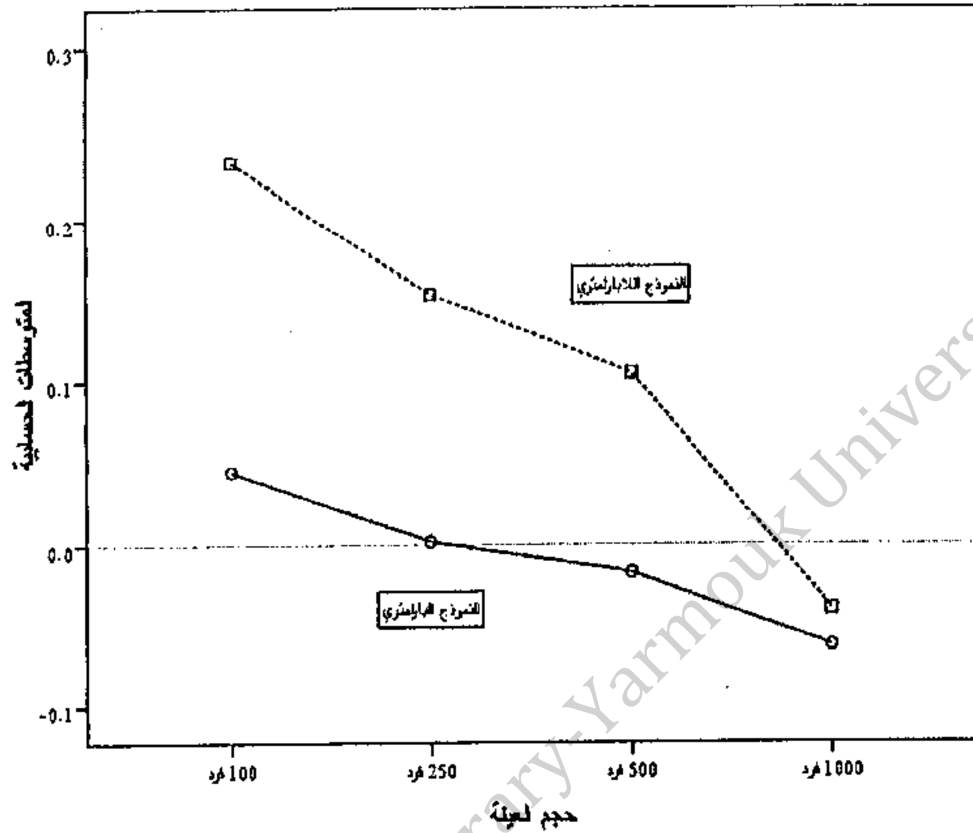


شكل 7: يوضح التفاعل الثنائي لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (طول الاختبار) على قيم مؤشر التحيز **BIAS** لتقديرات القدرة θ المقدرة

يلاحظ من الشكل 7 أنه تفاعل رتبي لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) عندما كانت أطوال الاختبارات (20، 60) فقرة لصالح النموذج البارامتري، مما يعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقديرات القدرة θ المقدرة مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى السبب إلى أن توزيع معالم فقرات الاختبارات الحقيقية (المولدة) هو

توزيع طبيعي (1، 0) لمعلمة الصعوبة وتوزيع منتظم لمعلمتي التمييز والتخمين وهو مناسب للنموذج البارامتري لذلك كانت تقديرات النموذج البارامتري أكثر دقة منه إلى النموذج اللابارامتري، إضافة إلى ذلك قد يعزى السبب إلى الطريقة المستخدمة في تقدير معلمة القدرة للنموذج البارامتري، حيث إن هذه الطريقة تعمل تكرار Iteration لخطوات التقدير للوصول إلى أعلى دقة في التقدير، وقد يعزى السبب إلى أن طريقة التقدير اللابارامترية لا تستطيع تقدير قدرات الأفراد الذين أجابوا على كل الفقرات إجابة صحيحة أو خطأ، مما أدى إلى تقدير قدرات متحيزة جدا لهؤلاء الأفراد. وعندما كان طول الاختبار 40 فقرة كانت الفروق بين النموذجين قليلة جدا، مما يعني أن النموذج اللابارامتري والبارامتري قد كان لهما نفس دقة تقديرات القدرة θ المقدرة.

وكذلك يتضح من الجدول 12 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، لابارامتري) مع متغير (حجم العينة)، والشكل 8 يوضح التفاعل الثنائي لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة.

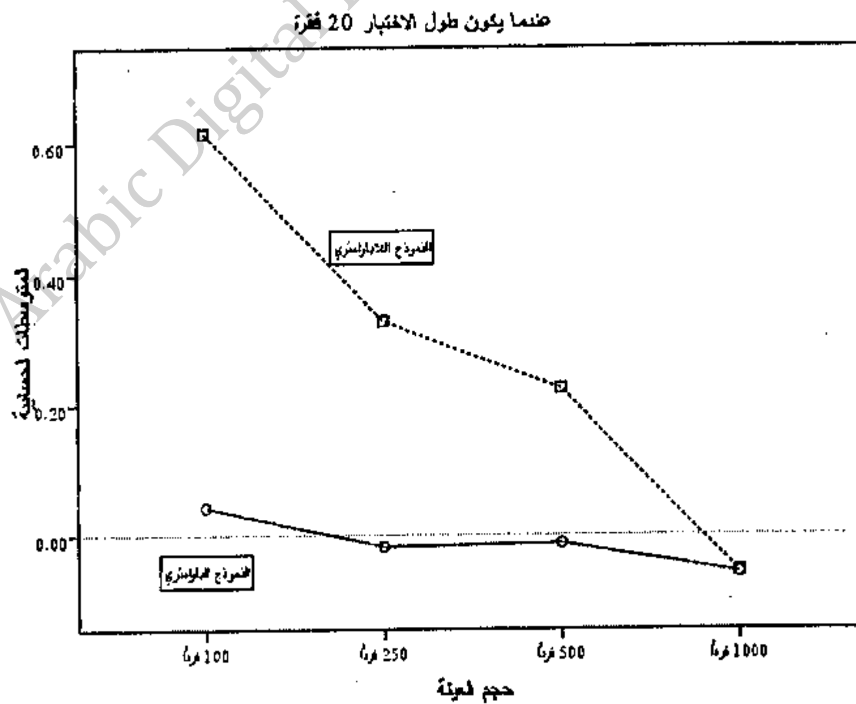


شكل 8: يوضح التفاعل الثنائي لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة

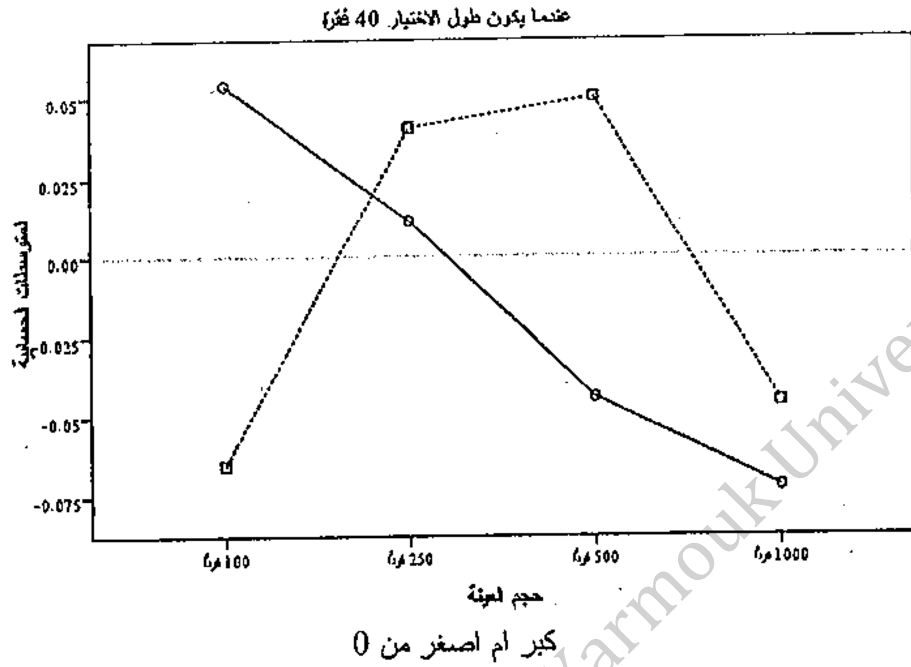
يلاحظ من الشكل 8 أنه تفاعل رتبي لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) عندما كانت حجوم العينات (100، 250، 500) فرد لصالح النموذج البارامتري، مما يعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقديرات القدرة θ المقدرة مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى السبب إلى أن توزيع معالم فقرات الاختبارات الحقيقة (المولدة) هو توزيع طبيعي (1، 0) لمعلمة الصعوبة وتوزيع منتظم لمعلمتي التمييز والتخمين، وهو مناسب للنموذج البارامتري، لذلك كانت تقديرات النموذج البارامتري أكثر دقة منه إلى النموذج اللابارامتري. إضافة إلى ذلك قد يعزى السبب إلى الطريقة المستخدمة في تقدير معلمة القدرة للنموذج البارامتري، حيث إن هذه الطريقة تعمل تكرار Iteration لخطوات

التقدير للوصول إلى أعلى دقة في التقدير، وعندما كان حجم العينة 1000 فرد لصالح النموذج اللابارامتري، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقديرات القدرة θ المقدرة مقارنةً بما كان عليه الأمر لدى النموذج البارامتري.

وأخيراً يتضح من الجدول 12 وجود فروق دالة إحصائيا عند مستوى الدلالة α=0.05 بين المتوسطات الحسابية لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة تعزى لتفاعل نوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة)، والأشكال (9، 10، 11) توضح التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغيري (طول الاختبار، وحجم العينة) على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة.

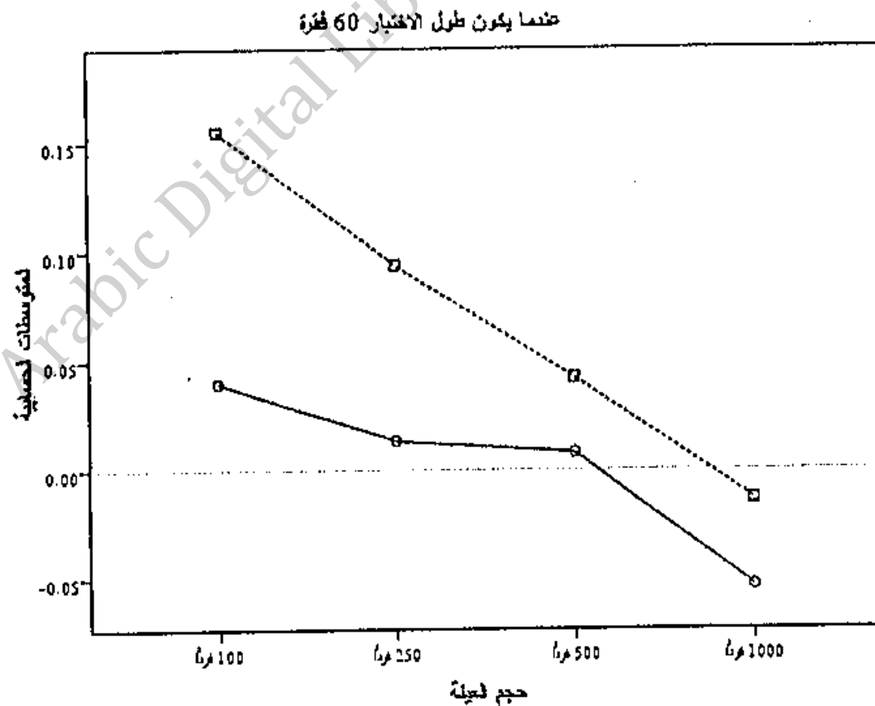


شكل 9: التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 20 فقرة على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة



كبر ام اصغر من 0

شكل 10التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 40 فقرة على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة



شكل 11: التفاعل لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 60 فقرة على قيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة

يلاحظ من الشكلين (9، 11) أنهما تفاعلان رتبيان لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة)

عندما يكون طول الاختبار (20، 60) فقرة، لصالح النموذج البارامتري، عندما تكون حجوم العينات (100، 250، 500) فرد، مما يُعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج البارامتري أقل ما يمكن، وقد يعزى السبب إلى طريقة التقدير والبرنامج المستخدم في النموذج اللابارامتري، وقد يعزى إلى شكل التوزيع لمعالم الفقرة الحقيقة (المولدة)، ولصالح النموذج اللابارامتري عندما يكون حجم العينة 1000 فرد، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج البارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج اللابارامتري أقل ما يمكن. كما يلاحظ من الشكل 10 أنه تفاعل لارتبي لقيم مؤشر التحيز BIAS لتقديرات القدرة θ المقدرة لنوع النموذج المستخدم (بارامتري، ولابارامتري) مع متغير (حجم العينة) عندما يكون طول الاختبار 40 فقرة، لصالح النموذج البارامتري عندما تكون حجوم العينات (100، 250، 500) فرد، مما يعني أن النموذج البارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج البارامتري أقل ما يمكن، ولصالح النموذج اللابارامتري عندما يكون حجم العينة 1000 فرد، مما يُعني أن النموذج اللابارامتري قد كان أكثر دقة جوهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج البارامتري، حيث جاء المتوسط الحسابي لمؤشر التحيز لتقديرات القدرة وفقاً للنموذج اللابارامتري أقل ما يمكن. وقد يعزى السبب إلى أن توزيع معالم فقرات الاختبارات الحقيقة (المولدة) هو توزيع طبيعي (1، 0) لمعلمة الصعوبة وتوزيع منتظم لمعلمتي التمييز والتخمين وهو مناسب للنموذج البارامتري، لذلك كانت

تقديرات النموذج البارامتري أكثر دقة منه إلى النموذج اللابارامتري. إضافة إلى ذلك قد يعزى السبب إلى الطريقة المستخدمة في تقدير معلمة القدرة للنموذج البارامتري، حيث إن هذه الطريقة تعمل تكرار Iteration لخطوات التقدير للوصول إلى أعلى دقة في التقدير.

وكذلك تم حساب قيم مؤشر RMSE كمؤشر من مؤشرات دقة التقدير لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 13.

جدول 13 قيم مؤشر RMSE لتقديرات القدرة θ المقدرة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

طول الاختبار	RMSE للقدرة باختلاف نوع النموذج:	حجم العينة 100 فرد	250 فرد	500 فرد	1000 فرد
20 فقرة	البارامتري	1.134	0.983	1.020	0.949
	اللابارامتري	1.795	1.232	1.227	1.117
40 فقرة	البارامتري	0.520	0.874	0.809	0.850
	اللابارامتري	1.282	0.945	0.944	0.887
60 فقرة	البارامتري	0.504	0.654	0.669	0.699
	اللابارامتري	1.089	0.838	0.730	0.606

يلاحظ من الجدول 13، أن قيم مؤشر RMSE لتقديرات القدرة حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم مؤشر RMSE لتقديرات القدرة حسب النموذج اللابارامتري في معظم حالات الدراسة، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج اللابارامتري، حيث جاء المتوسط الحسابي لمؤشر RMSE لتقديرات القدرة وفقاً للنموذج البارامتري أقل ما يمكن، باستثناء قيم مؤشر RMSE لتقديرات القدرة حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم مؤشر RMSE لتقديرات القدرة حسب النموذج اللابارامتري عندما يكون طول الاختبار 60 وحجم العينة 1000، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مما كان عليه الأمر لدى النموذج البارامتري، حيث جاء المتوسط الحسابي لمؤشر RMSE لتقديرات القدرة وفقاً للنموذج اللابارامتري أقل ما يمكن،

وقد يعزى السبب إلى الدالة الرياضية المستخدمة النموذج اللابارامتري، حيث إن هذه الدالة تقدر معلمة القدرة من منحنى خصائص الفقرة، الذي يتم تقديره من رتب الدرجات الخام، مما يؤدي إلى خسارة في كمية المعلومات، وبالتالي تقل دقة التقدير. وأخيراً، تم حساب قيم المساحات لدالة معلومات الاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 14.

جدول 14 مجموع قيم المساحات الخاصة بدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

طول الاختبار	الإحصائي	دالة المعلومات باختلاف النموذج:	حجم العينة 100 فرد	250 فرد	500 فرد	1000 فرد
20 فقرة	المتوسط الحسابي	اللابارامتري	6.75	2.99	1.87	1.50
		البارامتري	0.94	1.00	1.16	1.25
	الانحراف المعياري	اللابارامتري	4.90	2.24	1.66	1.15
		البارامتري	0.46	0.64	0.80	0.96
	المساحة	اللابارامتري	431.92	191.30	119.68	95.73
		البارامتري	60.39	64.23	74.30	79.82
40 فقرة	المتوسط الحسابي	اللابارامتري	1.48	1.13	1.09	0.77
		البارامتري	1.65	1.74	1.91	1.98
	الانحراف المعياري	اللابارامتري	1.31	1.06	0.93	0.67
		البارامتري	0.63	0.93	1.39	1.40
	المساحة	اللابارامتري	94.90	72.60	69.68	49.29
		البارامتري	105.57	111.25	121.99	126.81
60 فقرة	المتوسط الحسابي	اللابارامتري	1.51	0.56	0.46	0.38
		البارامتري	2.15	2.94	3.03	3.50
	الانحراف المعياري	اللابارامتري	1.37	0.49	0.43	0.34
		البارامتري	0.98	1.84	1.79	2.82
	المساحة	اللابارامتري	96.87	35.92	29.16	24.24
		البارامتري	137.55	187.95	193.77	224.09

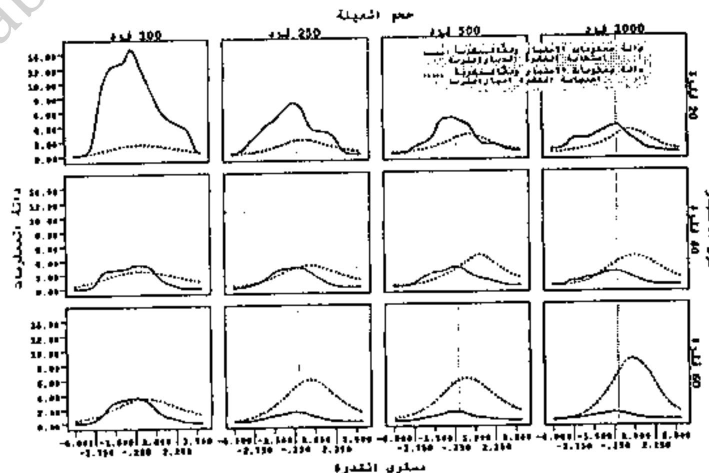
يلاحظ من الجدول 14، أن كافة قيم المساحة لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم المساحة لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 20 وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة بدلالة كِبَر مساحة دالة المعلومات لديه مما انعكس على انخفاض قيم الخطأ المعياري في تقدير القدرة مقارنةً بما كان عليه الأمر لدى

النموذج البارامتري، ويعزى السبب إلى أن النموذج اللابارامتري قد كان متحيزا في تقدير معالم الفقرات عند حجوم وأطوال العينات الصغيرة بشكل عام، ومعلمة التمييز (a) بشكل خاص، مما أدى إلى تحيز دالة المعلومات. في حين يلاحظ من الجدول 14، أن كافة قيم المساحة لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار) حسب النموذج البارامتري قد كانت أكبر ظاهرياً من قيم المساحة لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار (40، 60)، وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة بدلالة كِبَر مساحة دالة المعلومات لديه مما انعكس على انخفاض قيم الخطأ المعياري في تقدير القدرة مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى السبب إلى كبر حجم العينة، حيث إن دالة معلومات الاختبار تزداد بزيادة عدد الفقرات.

كما يلاحظ من الجدول 14، أن كافة قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة وطول الاختبار) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 20، وحجم العينة (100، 250، 500، 1000). بالمقابل يلاحظ من الجدول 14 أن قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري)، وباختلاف متغيري (حجم العينة، وطول الاختبار) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 40 وحجم العينة (100، 250)، في حين كانت قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، لابارامتري) وباختلاف متغيري (حجم العينة وطول الاختبار) حسب

النموذج البارامتري أكبر ظاهرياً من قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 40 وحجم العينة (500، 1000). وأخيراً، يلاحظ من الجدول 14، أن قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري)، وباختلاف متغيري (حجم العينة، وطول الاختبار) حسب النموذج البارامتري قد كانت **أصغر** ظاهرياً من قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 60 وحجم العينة 100، في حين كانت قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار) حسب النموذج البارامتري أكبر ظاهرياً من قيم الانحراف المعياري لدالة المعلومات للاختبار ذي البيانات المولّدة حسب النموذج اللابارامتري عندما يكون طول الاختبار 60 وحجم العينة (250، 500، 1000).

والشكل 12، يبين دالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار).



شكل 12: دالة المعلومات للاختبار ذي البيانات المولّدة باختلاف نوع النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار).

كما تم حساب المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار ذي البيانات المولّدة إلى دالة المعلومات الحقيقية للاختبار ذي البيانات المُولّدة باختلاف النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)، وذلك كما هو مبين في الجدول 15.

الجدول 15 المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار إلى دالة المعلومات الحقيقية للاختبار باختلاف النموذج المستخدم (بارامتري، ولابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)

حجم العينة	الكفاءة النسبية لدالة معلومات الاختبار باختلاف النموذج المستخدم:	طول الاختبار 20 فقرة	40 فقرة	60 فقرة
100	اللابارامتري/الحقيقي	9.552	1.784	0.872
فرد	البارامتري/الحقيقي	1.585	2.434	1.507
250	اللابارامتري/الحقيقي	4.724	1.203	0.370
فرد	البارامتري/الحقيقي	1.361	1.838	1.534
500	اللابارامتري/الحقيقي	2.257	1.140	0.231
فرد	البارامتري/الحقيقي	1.546	1.489	1.747
1000	اللابارامتري/الحقيقي	3.179	0.986	0.254
فرد	البارامتري/الحقيقي	1.445	1.499	1.422

يلاحظ من الجدول 15، أن كافة قيم المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار ذي البيانات المولّدة وفقاً للنموذج اللابارامتري إلى دالة المعلومات الحقيقية للاختبار ذي البيانات المولدة قد كانت **أصغر** ظاهرياً من كافة قيم المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار ذي البيانات المولّدة وفقاً للنموذج البارامتري إلى دالة المعلومات الحقيقية للاختبار ذي البيانات المولدة عندما يكون طول الاختبار (40، 60) فقرة، وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر كفاءة ظاهرياً في حساب دالة المعلومات الاختبار مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، كما يلاحظ من الجدول 15، أن كافة قيم المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار ذي البيانات المولّدة وفقاً للنموذج البارامتري إلى دالة المعلومات الحقيقية للاختبار ذي البيانات المولدة قد كانت أصغر ظاهرياً من كافة قيم المتوسطات الحسابية للكفاءة النسبية الخاصة بدالة المعلومات المقدرة للاختبار ذي البيانات المولّدة وفقاً للنموذج اللابارامتري إلى دالة المعلومات الحقيقية للاختبار ذي البيانات المولدة عندما يكون طول الاختبار 20 فقرة وحجم العينة

(100، 250، 500، 1000)، مما يعني أن النموذج اللابارامتري قد كان أكثر كفاءة ظاهرياً في حساب دالة المعلومات الاختبار مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، مما يعني أن النموذج اللابارامتري أكثر كفاءة من النموذج البارامتري عند أطوال الاختبارات الصغيرة، والنموذج البارامتري أكثر كفاءة عند أطوال الاختبارات الكبيرة.

ثالثاً. النتائج المتعلقة بسؤال الدراسة الثالث: "ما مدى التوافق في تقدير معالم الفقرة والقدرة باختلاف النموذج (بارامتري، واللابارامتري) وباختلاف متغيري (حجم العينة، وطول الاختبار)؟"

وللإجابة عن سؤال الدراسة؛ تم حساب معاملات الارتباط البينية لقيم تقديرات القدرة θ المقدرة باختلاف نوع النموذج (بارامتري، ولابارامتري) والحقيقية عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250، 500، 1000)، وذلك كما هو مبين في الجدول 16.

جدول 15 قيم معاملات الارتباط البينية لقيم تقديرات القدرة θ المقدرة باختلاف نوع النموذج (بارامتري، ولابارامتري) والحقيقية وباختلاف متغيري (طول الاختبار، وحجم العينة)

حجم العينة	النموذج	طول الاختبار: 20 فقرة		40 فقرة		60 فقرة	
		الحقيقية	البارامترية	الحقيقية	البارامترية	الحقيقية	البارامترية
100 فرد	الحقيقية						
	البارامترية	0.80		0.77		0.86	
	اللابارامترية	0.72	0.83	0.76	0.90	0.81	0.92
250 فرد	الحقيقية						
	البارامترية	0.77		0.79		0.88	
	اللابارامترية	0.68	0.88	0.73	0.93	0.86	0.96
500 فرد	الحقيقية						
	البارامترية	0.74		0.80		0.88	
	اللابارامترية	0.70	0.86	0.77	0.90	0.86	0.95
1000 فرد	الحقيقية						
	البارامترية	0.78		0.82		0.885	
	اللابارامترية	0.74	0.88	0.79	0.91	0.880	0.94

يلاحظ من الجدول 16، أن مدى التوافق كمؤشر ثقة بين قيم تقديرات القدرة θ المقدرة حسب النموذج البارامتري وبين قيم القدرة θ الحقيقية قد كان أكبر ظاهرياً من مدى التوافق كمؤشر ثقة بين قيم تقديرات القدرة θ المقدرة حسب النموذج اللابارامتري، وقيم القدرة θ الحقيقية عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250،

1000، 500)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة القدرة مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري.

كما يلاحظ من الجدول 16، أن مدى التوافق بين قيم تقديرات القدرة θ المقدرة حسب النموذج البارامتري، وقيم تقديرات القدرة θ المقدرة حسب النموذج اللابارامتري قد صُنِّفت ضمن درجة (كبيرة) وفقاً لمعيار (Hinkle, Wiersma & Jurs, 1988) عندما يكون طول الاختبار 20 وحجم العينة (100، 250، 500، 1000)K مما يعني أن النموذج البارامتري يتكافأ نوعاً ما في تقدير معلمة القدرة مع النموذج اللابارامتري. في حين يلاحظ من الجدول 16، أن مدى التوافق بين قيم تقديرات القدرة θ المقدرة حسب النموذج البارامتري، وقيم تقديرات القدرة θ المقدرة حسب النموذج اللابارامتري قد صُنِّفت ضمن درجة (كبيرة جداً) عندما يكون طول الاختبار (40، 60)، وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري يتكافأ في تقدير معلمة القدرة مع النموذج اللابارامتري.

كما تم حساب معاملات الارتباط البينية لقيم تقديرات معالم الفقرات (a, b, c) المقدرة باختلاف نوع النموذج (بارامتري، ولابارامتري)، والحقيقية وباختلاف متغيري (طول الاختبار، وحجم العينة)، وذلك كما هو مبين في الجدول 16.

جدول 16 قيم معاملات الارتباط البينية لقيم تقديرات معالم الفقرات (a, b, c) المقدرة باختلاف نوع النموذج (بارامتري، ولابارامتري) والحقيقية وباختلاف متغيري (طول الاختبار، وحجم العينة)

طول الاختبار	حجم العينة	نظرية استجابة الفقرة:	الإحصائي	معلمة التمييز		معلمة الصعوبة		معلمة التخمين	
				الحقيقية	البارامترية	الحقيقية	البارامترية	الحقيقية	البارامترية
20 فقرة	100 فردا	البارامترية	معامل الارتباط	0.38		0.95		0.00	
			الدلالة الإحصائية	0.098		0.000		0.994	
		اللابارامترية	معامل الارتباط	0.30	0.23	0.24	0.29	0.20	0.33
			الدلالة الإحصائية	0.205	0.319	0.313	0.211	0.398	0.158
	250 فردا	البارامترية	معامل الارتباط	0.48		0.78		0.29	
			الدلالة الإحصائية	0.030		0.000		0 220	
		اللابارامترية	معامل الارتباط	-0.37	-0.49	-0.03	-0.04	0.32	0.69
			الدلالة الإحصائية	0.114	0.027	0.912	0.883	0.169	0.006
	500 فردا	البارامترية	معامل الارتباط	0.69		0.97		0.16	
			الدلالة الإحصائية	0.001		0.000		0.501	
		اللابارامترية	معامل الارتباط	0.62	0.72	-0.14	-0.14	-0.06	0.74
			الدلالة الإحصائية	0.004	0.000	0.564	0.547	0.792	0.000
	1000 فردا	البارامترية	معامل الارتباط	0.63		0.92		0.40	
			الدلالة الإحصائية	0.003		0.000		0.083	
		اللابارامترية	معامل الارتباط	0.18	0.25	0.34	0.43	0.16	0.46
			الدلالة الإحصائية	0.450	0.263	0.143	0.058	0.611	0.039
40 فقرة	100 فردا	البارامترية	معامل الارتباط	0.29		0.88		0.46	
			الدلالة الإحصائية	0.072		0.000		0.003	
		اللابارامترية	معامل الارتباط	0.04	0.03	-0.19	-0.34	0.27	0.65
			الدلالة الإحصائية	0.794	0.864	0.252	0.031	0.088	0.000
	250 فردا	البارامترية	معامل الارتباط	0.60		0.94		0.273	
			الدلالة الإحصائية	0.000		0.000		0.068	
		اللابارامترية	معامل الارتباط	0.07	0.02	0.17	0.13	0.27	0.73
			الدلالة الإحصائية	0.647	0.900	0.308	0.414	0.095	0.000
	500 فردا	البارامترية	معامل الارتباط	0.61		0.91		0.39	
			الدلالة الإحصائية	0.000		0.000		0.013	
		اللابارامترية	معامل الارتباط	0.05	0.25	0.63	0.65	0.21	0.59
			الدلالة الإحصائية	0.738	0.121	0.000	0.000	0.189	0.000

طول الاختبار	حجم العينة	نظرية استجابة الفقرة	الإحصائي	معلمة التمييز: الحقيقية	معلمة التمييز: البارامترية	معلمة الصعوبة: الحقيقية	معلمة الصعوبة: البارامترية	معلمة التخمين: الحقيقية	معلمة التخمين: البارامترية
	1000 فردا	البارامترية	معامل الارتباط	0.86		0.94		0.37	
			الدلالة الإحصائية	0.000		0.000		0.017	
		اللابارامترية	معامل الارتباط	-0.06	0.08	0.56	0.68	0.33	0.71
			الدلالة الإحصائية	0.705	0.643	0.000	0.000	0.039	0.000
60 فقرة	100 فردا	البارامترية	معامل الارتباط	0.41		0.91		0.08	
			الدلالة الإحصائية	0.001		0.000		0.544	
		اللابارامترية	معامل الارتباط	-0.16	-0.07	0.10	0.04	-0.03	0.47
			الدلالة الإحصائية	0.228	0.618	0.458	0.745	0.817	0.000
	250 فردا	البارامترية	معامل الارتباط	0.68		0.92		0.15	
			الدلالة الإحصائية	0.000		0.000		0.239	
		اللابارامترية	معامل الارتباط	0.02	0.03	0.30	0.42	-0.05	0.69
			الدلالة الإحصائية	0.903	0.792	0.020	0.001	0.680	0.000
	500 فردا	البارامترية	معامل الارتباط	0.77		0.96		0.21	
			الدلالة الإحصائية	0.000		0.000		0.107	
		اللابارامترية	معامل الارتباط	0.01	0.03	0.01	0.01	0.15	0.61
			الدلالة الإحصائية	0.969	0.814	0.948	0.937	0.247	0.000
	1000 فردا	البارامترية	معامل الارتباط	0.72		0.95		0.27	
			الدلالة الإحصائية	0.000		0.000		0.037	
		اللابارامترية	معامل الارتباط	0.22	0.56	0.54	0.61	0.07	0.41
			الدلالة الإحصائية	0.088	0.000	0.000	0.000	0.692	0.001

يلاحظ من الجدول 16، أن النتائج الخاصة به قد كانت على النحو الآتي:

أ. **فيما يخص معلمة التمييز:** إن مدى التوافق كمؤشر ثقة بين قيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) الحقيقية قد كان أكبر ظاهرياً من مدى التوافق بين قيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) المقدرة حسب النموذج اللابارامتري، وقيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) الحقيقية عندما يكون طول الاختبار (20، 40، 60) وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة الفقرات (a) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى ذلك إلى أن النموذج اللابارامتري كان متحيزا في تقدير معلمة التمييز (a).

كما يلاحظ من الجدول 17، أن مدى التوافق بين قيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة التمييز لفقرات الاختبار ذي البيانات المُوَّلدة (a) المقدرة حسب النموذج اللابارامتري قد كان

دالاً إحصائيا عند مستوى الدلالة α=0.05 عندما يكون طول الاختبار 20 فقرة وحجم العينة (250، 500) فرد، وعندما يكون طول الاختبار 60 فقرة وحجم العينة 1000 فرد.

ب. **فيما يخص معلمة الصعوبة:** إن مدى التوافق كمؤشر ثقة بين قيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) الحقيقية قد كان أكبر ظاهرياً من مدى التوافق بين قيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) المقدرة حسب النموذج اللابارامتري، وقيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) الحقيقية عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة الفقرات (b) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، وقد يعزى ذلك الى ان النموذج اللابارامتري لا يستطيع تقدير معلمة صعوبة الفقرات التي اجاب عليها او لم يجيب عليها كل الأفراد، وبالتالي تم تقدير معلمة الصعوبة لهذه الفقرات بتحيز كبير سواء للأعلى او للأسفل.

كما يلاحظ من الجدول 17، أن مدى التوافق بين قيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة الصعوبة لفقرات الاختبار ذي البيانات المُوَلّدة (b) المقدرة حسب النموذج اللابارامتري قد كان دالاً إحصائيا عند مستوى الدلالة α=0.05 عندما يكون طول الاختبار 40 فقرة وحجم العينة (100، 500، 1000) فرد، وعندما يكون طول الاختبار 60 فقرة وحجم العينة (250، 1000) فرد.

ج. **فيما يخص معلمة التخمين: إن مدى التوافق كمؤشر ثقة بين قيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c)** الحقيقية قد كان أكبر **ظاهرياً من** مدى التوافق بين قيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج اللابارامتري، وقيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) الحقيقية، عندما يكون طول الاختبار (20، 40، 60)، وحجم العينة (100، 250، 500، 1000)، مما يعني أن النموذج البارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة الفقرات (c) مقارنةً بما كان عليه الأمر لدى النموذج اللابارامتري، باستثناء أن مدى التوافق كمؤشر ثقة بين قيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) الحقيقية قد كان **أصغر** ظاهرياً من مدى التوافق بين قيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج اللابارامتري، وقيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) الحقيقية عندما يكون طول الاختبار 20 وحجم العينة 100، مما يعني أن النموذج اللابارامتري قد كان أكثر دقة ظاهرياً في تقدير معلمة الفقرات (c) مقارنةً بما كان عليه الأمر لدى النموذج البارامتري، يمكن الاستدلال على ان النموذج اللابارامتري لا يتأثر في صغر حجم العينة وطول الاختبار في تقدير معلمة الفقرات (c)، وقد يعزى ذلك إلى أن طريقة التقدير في النموذج اللابارامتري لا تتأثر في شكل التوزيع للمعالم الحقيقية؛ وذلك لأنها تعتمد الرتب في تقدير منحنى خصائص الفقرة، ومن ثم تقدر المعالم.

كما يلاحظ من الجدول 17، أن مدى التوافق بين قيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج البارامتري، وقيم تقديرات معلمة التخمين لفقرات الاختبار ذي البيانات المُوّلدة (c) المقدرة حسب النموذج اللابارامتري قد كان

دالاً إحصائيًا عند مستوى الدلالة $\alpha=0.05$ عندما يكون طول الاختبار 20 فقرة، وحجم العينة (250، 500، 1000) فرد، وعندما يكون طول الاختبار (40، 60) فقرة، وحجم العينة (100، 250، 500، 1000) فرد.

الخلاصة والتوصيات

أظهرت نتائج الدراسة أن طريقة تقدير معالم الفقرة والقدرة باستخدام طريقة MML البارامترية قد كانت أكثر دقة من طريقة KS اللابارامترية باختلاف متغيري (حجم العينة، وطول الاختبار)، في حالات الدراسة جميعها تقريبا، وقد يعزى ذلك إلى أسباب عدة.

- الدالة الرياضية المستخدمة في الطريقة اللابارامتري KS، حيث إن هذه الطريقة تعتمد الرتب في تقدير منحنى خصائص الفقرات، وتعطي القيم المتساوية رتباً بشكل عشوائي، وبعد ذلك يتم تقدير المعالم، ولذلك فإن طريقة KS لا تفترض توزيعات معينة للبيانات الحقيقية (المولدة)، وبالتالي فإن هذه الطريقة تحتاج إلى حجوم عينات وأطوال اختبار أقل من طريقة MML البارامترية. ولكن عملية تحويل البيانات من مستوى القياس الفئوي إلى مستوى القياس الرتبي يؤدي إلى خسارة في كمية المعلومات (تقل كمية المعلومات)، وكنتيجة لذلك فإن دقة تقدير معالم الفقرة والقدرة بطريقة KS أقل من دقة تقدير معالم الفقرة والقدرة باستخدام طريقة MML البارامترية، ولكن هذه الطريقة تعطي تقديرات اكثر دقة عند حجوم العينات واطوال الاختبارات الصغيرة .
- البرامج الحاسوبية المستخدمة.

البرنامج المستخدم لتقدير معالم الفقرة والقدرة بالطريقة البارامترية MML هو Bilog–Mg، وهذا البرنامج يعمل اعادة تدوير Iteration لتقدير المعالم للوصول إلى محك معين، قد تصل إلى 100 مرة أو أكثر للوصول بالنتائج إلى أعلى دقة، أما البرنامج المستخدم لتقدير معالم الفقرة والقدرة بالطريقة اللابارامترية KS هو Test Graf، وهذا البرنامج يقدر المعالم مباشرة دون تدوير لعملية التقدير، ولذلك فإن دقة تقدير معالم الفقرة والقدرة بالطريقة البارامترية أعلى منه في الطريقة اللابارامترية.

- توزيعات معالم الفقرة والقدرة الحقيقية المستخدمة في هذه الدراسة.

إن توزيع معالم الفقرة والقدرة الحقيقية المستخدمة في هذه الدراسة هي توزيعات طبيعية لمعلمتي القدرة والصعوبة بواقع متوسط حسابي مقداره صفر، وانحراف معياري مقداره 1، والتوزيع المنتظم لمعلمتي التمييز والتخمين، وهذه التوزيعات لمعالم الفقرة والقدرة الحقيقية أنسب لطريقة التقدير البارامترية منه إلى طريقة التقدير اللابارامترية، بالرغم من صغر حجم العينة وصغر طول الاختبار في بعض حالات الدراسة.

وفي ظل جميع الفروق في الثقة والكفاءة والدقة في التقدير بين طريقة التقدير البارامترية MML والطريقة اللابارامترية KS في تقدير معالم الفقرات ومعلمة القدرة، التي كانت في معظم الأحيان لصالح الطريقة البارامترية، فإنه يمكن تعويض ذلك في أن طريقة KS تصلح لتقدير معالم الفقرات ومعلمة القدرة للبيانات الرتبية، وأنها لا تفترض أي شكل لتوزيع البيانات الخام، وذلك أنها تحول الدرجات الملاحظة إلى رتب، وبالتالي فإنها تصلح للبيانات الرتبية والفئوية وأنها أسرع بـــ 500 مرة من الطريقة البارامترية.

ويوصي الباحث باستخدام طريقتي التقدير البارامترية واللابارامترية في تقدير معالم الفقرة والقدرة واستخدام الطريقة اللابارامترية في حال كانت البيانات رتبية في طبيعتها، أو عندما يكون حجم العينة وطول الاختبار لا تتناسب مع الطريقة البارامترية.

ولم تتمكن الدراسة من تقصي دقة التقدير في ظل بيانات حقيقية، وتغير حجوم العينات وأطوال اختبارات أكثر تباين، واختلاف توزيع القدرات واختلاف توزيع معالم الفقرات، واختلاف النموذج، وعليه يوصي الباحث تقصي دقة تقدير معالم الفقرة والقدرة باختلاف متغيري (حجم العينة، وطول الاختبار) في ظل ما يأتي:

- حجم العينة، حيث أمكن للباحث تقصي دقة التقدير بعدد يتراوح ما بين 100 فرد إلى 1000 فرد، وبالتالي يمكن دراسة أثر حجوم عينات أكثر تباين من ذلك على دقة التقدير باختلاف النموذجين.
- أطوال الاختبار، حيث أمكن للباحث تقصي دقة التقدير بأطوال اختبار تتراوح ما بين 20 فقرة إلى 60 فقرة، وبالتالي يمكن دراسة أثر اطول اختبارات أكثر تباين من ذلك على دقة التقدير باختلاف النموذجين.
- تقصي أثر باراميتر التهذيب (Smoothing Parameter) حيث أمكن للباحث تقصي دقة التقدير لبراميتر التهذيب المستخدم في برنامج TESTGRAF، ويمكن تقصي أثر اختلاف قيمة باراميتر التهذيب (Smoothing Parameter) على دقة تقدير معالم الفقرة والقدرة في النموذج اللابارامتري.
- بيانات حقيقية، حيث أمكن للباحث تقصي دقة التقدير باستخدام بيانات مولدة، وبالتالي يمكن تقصي دقة التقدير وفق بيانات حقيقية.
- توزيع القدرات للمفحوصين، حيث تمكن الباحث من تقصي دقة التقدير في ظل توزيع طبيعي وسطه صفر، وانحراف معياري واحد، وبالتالي يمكن دراسة أثر توزيع القدرات على دقة التقدير باختلاف النموذج المستخدم في التقدير.
- توزيع معالم الفقرات حيث أمكن للباحث تقصي دقة التقدير في ظل توزيع طبيعي لمعلمة الصعوبة وسطه صفر، وانحرافه المعياري واحد، وتوزيع منتظم لمعلمتي التمييز والتخمين، وبالتالي يمكن دراسة أثر التوزيع (توزيعات ملتوية) لمعالم الفقرات على دقة التقدير باختلاف الطريقة المستخدمة في التقدير.

- ثم في هذه الطريقة تقصي دقة تقدير معالم الفقرة، والقدرة باستخدام طريقة MML البارامترية وطريقة KS اللابارامترية، وبالتالي يمكن تقصي دقة تقدير معالم الفقرة والقدرة باستخدام طرق أخرى.
- استخدام النموذج ثنائي المعلم والنموذج أحادي المعلم لتقصي دقة تقدير معالم الفقرة والقدرة باختلاف النموذج المستخدم (بارامتري، واللابارامتري)، وذلك للسيطرة على معلمة التخمين .

المراجع

المراجع العربية

الدردير،عبد المنعم احمد. (2006). **الإحصاء البارامتري واللابارامتري في اختبار فروض البحوث النفسية والتربوية والاجتماعية**. القاهرة: عالم الكتب .

الشريفين، نضال كمال.(2012) .أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. **المجلة التربوية، 26**(104)،177-238.

علام، صلاح الدين محمود.(1993). **الأساليب الإحصائية الاستدلالية البارامترية واللابارامترية في تحليل بيانات البحوث النفسية والتربوية**. القاهرة: دار الفكر العربي.

علام، صلاح الدين محمود.(2005). **نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي**. القاهرة: دار الفكر العربي.

علام، صلاح الدين. (2000). **القياس والتقويم التربوي النفسي، أساسياته وتطبيقاته وتوجهاته المعاصرة**. القاهرة: دار الفكر العربي.

عودة، احمد سليمان. (2010). **القياس والتقويم في العملية التدريسية**.اربد : دار الأمل للنشر والتوزيع.

عودة، احمد سليمان. والخليلي، خليل يوسف. (2000). **الإحصاء للباحث في التربية والعلوم الإنسانية**. اربد: دار الأمل للنشر والتوزيع.

النبهان، موسى. (2004). **أساسيات القياس في العلوم السلوكية**. عمان: دار الشرق للنشر والتوزيع.

المراجع الأجنبية

Baker.(2001). **The Basics of Item Response Theory** (2nd Ed.). ERIC Clearinghouse on Assessment & Evaluation: USA.

Birnbaum, A. (1986). **Some latent trait Models and their Use in Inferring an Examinee's Ability.** In F.M. Lord & M. R. Novick (Eds). Statistical theories of mental test scores. Reading. MA: Addison- Wesley.

Cliff, N., & Keats , J . (2003). **Ordinal Measurement in the Behavioral Sciences .** LawrenceElbaum Associates, NJ: Mahwan.

Crocker, L., &Algina, J. (1986). **Introduction to classical and modern test theory.** NY: Holt, Rinhart&Winston.

Crocker, L., &Algina, J. (1987). **Introduction to Classical and Modern Test Theory**. New York: Harcourt Jovanovich.

De- Ayala, R. (1993). Computerized Adaptive Testing Using The Partial Credit Model : Effects on Item Pool Characteristics and Different Stopping Rules. **Educational and Psychological Measurement, 53**(3), 61-77.

De la Torre, Jimmy; Yuan Hong (2010). Parameter Estimation With Small Sample Size A Higher-Order IRT Model Approach.**Applied Psychological Measurement**.. 34 (4), p267-285.

Degruijter, N., & Van der Kamp, J. (2005). **Statistical test theory for education and Psychology**. Retrieved in 2013 from: www.leidenuniv.nl/griuijterdnme.

Fitzpatrick,& Ann . R. (2001) . The effects of test length and sample Size on the reliability and equating of tests Composed of constructed – response item .**Applied Measurement in Education,** 14(1): 412-425.

Fu, Qiong (2010) **Comparing Accuracyof ParameterEstimationUsing IRT Models in the Presence of Guessing** .. Unpublished PhD dissertation, Illinois University, USA.

Gao, F., & Chen, L. (2005). Bayesian or Non-Bayesian: A Comparison study of Item Parameter Estimation in the three – Parameter Logistic Model. **Applied Measurement in Education, 18** (4), 351 – 380.

Ghisiell,E.(1981). **Measurement Theory For the Behavioral Sciences** Sanfrancisco; W .H . Freeman and company.

Gu ,H., and sinharay, S: (2011). **MeasurmentErrorin Nonparametric Item Response Curve Estimation. Research Report**, ETS, RR-11-28. U.S.A.

Hambelton, K.(1993). Principles and selected applications of item response theory .in R. L.linn(Ed) **Educational Measurement.**(3 ed.)(pp.147-200)phoenix:theory press.

Hambleton, K., Swaminathan, H., & Rogers, H. (1991). **Fundamental of Item Response Theory.** Newbury Park, CA: sage publications, Inc.

Hambleton, R. (1989) . **Principles and Selected Applications of Item response theory.** In Robert L. Linn, (Ed), **Educational Measurement** (3rd ed.). New York : American Council on Education, Macmillan Publishing Company.

Hambleton, R.,. &Swaminthan, H. (1985). **Item response Theory Principle and Application.** Boston: Kluwer: Nijhoff Publishing.

Han, T., &Hambelton,. K (2007). **User Manual for WinGen: Windows Software That Generates IRT Model Parameters and Item Response.** Holt Rinehart and Wiston, New York.

Hansen, A. (2004). **Predicting The Distribution Of A goodness Of Fit Statistics Appropriate For Use With Performance – Based Assessments** (Doctoral dissertation, university of Pittsburgh, 2004 .

Harwell , M .et al. (1996) Monecarlo Studies . in Item Response Theory. **Applied Psychological Measurement,20** (2): 22-56.

Johnson, M. (2006). **Nonporamctuic Estimation of Item and Respond ant locations from Unfolding type items.** NY: the psychometric society.

Junker ,B ., &Sijtsma, K .(2001). Cognitive Assessment Models With Few Assumptions & Connections With Nonparametric Item Response Theory, **Applied Psychological Measurement,25**(4), 258-272.

Kingma, J ., &Tenvergert , E. (1985). A Nonparametric Scale Analysis of development of Conservation. **Applied Psychological Measurement, 9**(2),375-387.

Koning, E.; Sijtsma, K.&Hamers, J.(2002). Comparison of Four IRT Models When Analysing Two Test for Inductive Reasoning **.Applied Psychological Measurement, 26**(3),302-320

Lew,P., Dunbar, S., and Kolen, M. (2004). A Compression of parametric Approaches to Item Analysis for multiple – choice tests, **Education and psychological measument, G4** (4), 565-587

Liang, T .(2010).An Assessment of the Nonparametric Approach of Evaluating the fit of Item Response Model. **Dissertation Abstract International,** (UMI No. 3397726).

Linden, W. &Hambleton, R. (1997). **Handbook of Modern Item Response Theory.** Springer-Verlag. New York Inc.: New YorkBerlinHeidelberg .

Lord, F .(1977). Practical Application of Item Characteristic Curve Theory .**Journal of Educational Measurement, 14**,(1), 117-138.

Lord, F. M. (1980). **Application of item response theory to practical testing problems**. Hillsadle, New Jersey: Lawrence Erlbaum Associates.

Mckinley, R. L., and Mills, C. N. (1985). A comparison of several goodness –of- fit statistics. **Applied psychological measurement 9** (1), 49-57.

Meijer, R.and Baneke, J. (2004). Analyzing Psychopathology Items: A Case for Nonparametric Item Response Theory Modeling. **American Psychological Association,** 9(3), 354–368.

Mislevey, R., & Bock, R. D. (2003). **BILOG (Version 3) (Computer software),**Lincolnwood, IL: Scientific Software International.

Mislevy, R. (1986). Bayes modal estimation in item response models.**Psychometrika, 51**(4), 177 – 195.

Mokken, J., & Lewis, C. (1982). A Nonparametric Approach to the Analysis of Dichotomous Item Response. **Applied Psychological Measurement**, 6, (2),417-430.

Orlando, M., and Thissen, D. (2000). Likelihood- based item- fit indices for dichotomous item response theory models .**Applied** . **Psychological Measurement**, 24(1), 50-64.

Ramsay,J. (1991). Kernel Smoothing Approaches to Non-para Metric Item characteristic Curve Estimation. **Psychometrika, 56** (4), 611 – 630.

Ramsey, J.(2000). **Test Graf: A program For the graphical analysis of multiple-choice tests and questionnaire data [Computer Software and manual]** .Retrieved from http://www.psych.mcgill.ca/faculty/ramsy.htm. Reeve, B. (2004). **Applications of item response theory (IRT) modeling for building an evaluating questionnaires measuring patient-reported outcomes**. Web site:
httb://outcomes.cancer.gov/conference/irt/ree-ve.pdf.

Reckase, &Mark, D. (1978) A Comparison of the one and three Parameter logistic model for item calibration. **Paper Presented at the annual meeting of the American Educational Research Association,** Toronto ,Canada. {on - Line}Available : http :// eric .ed .gov.

Schmitt, T.; Sass, D.; Sullivan, D.; Walker, C. (2010).A Monte Carlo Simulation Investigating the Validity and Reliability of Ability Estimation in Item Response Theory with Speeded Computer Adaptive Tests. **International Journal of Testing,** 10: 230–261,

Sijtsma, K .&Hamker, B . T .(2000). A taxonomy of IRT models for ordering of persons and item using simple sum scores **.Journal of Educational and Behavioral Statistics, 25**(2), 391-415 .

Sijtsma, K .(1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. **Applied Psychological Measurement, 22,3-31.**

Sijtsma, K. &Molenaar, I. (2002). **Introduction to Nonparametric Item Response Theory.** Sage Publication, International Educational and Professional Publisher. Thousand Oaks : London . New Delhi.

Sijtsma, K.; Emons, W. H. M.; Boumeester, S.; Nyklicek, I. &Rorda, L. D. (2007). Nonparametric IRT analysis of quality of life scales and its application to the world health organization quality of life scale (WHOQOL-bref). **Quality of Life Research,17**(3), 275-290.

Trevisan, M. S., Sax, g. & Michael, W. B. (1991). The effect of number of options per item and student ability on test validity and reliability. **Educational and Psychological Measurement, 51**(4), 829 – 837.

Van der Linden,W.J,&Hambleton , R .K.(1997) .Part IV :Non-Parametric models introduction in W. J. van der Linden & R. K. Hambleton(Eds),Handbook of modern item response theory (pp.347-349).New York :Springer- Verlag .

Wagner, T.A, &Harvey ,R . J .(2003) . **Developing a new Critical thinking test using item response theory.**Paper Presented at the 2003 annual Conference of the Society for industrial and organizational Psychology . orlando

Wang, T ., &Vispoel, W . P . (1998). Properties of ability Estimation Methods in Computerized Adaptive Testing. **Journal of Educational Measurement, 35**(2), 109 – 135 .

Wang, Xiang-Bo; Harris, Vincent ; Roussos, Louis (2003). **Effects of Multidimensionality on IRT Item Characteristics and True Score Estimates: Implications for Computerized Test Assembly. Computerized Testing Report. LSAC Research Report Series.**Law School Admission Council, Newtown, PA.

Weiss,D.J. (2009) **.Termination Criteria in Computerized Adaptive Tests : Variable – Length CATs Are Not Biased.** Paper Presented in the 2007 GMAC Conference on Computerized Adaptive Testing .Minneapolis.

Zeng, L. (1997). Implementation of marginal Bayesian estimation with four – parameter beta prior distributions. **Applied Psychological Measurement, 21(2)**, 143 – 156.

Zickar, M .(1997) . Identifying Untraited Individuals Using Model-based Measurement. **Dissertation Abstract International.**(UMI NO.9737304).

Zimowski, M. F., Muraki, E., Mislevy, R., J., & Bock, R. D. (2003). **Biilog, Mg3 (Computer software). In M. du Toit (ed), Irt from SSI: Bilog MG, Multiple, Parscale, test fact.** Lincolnwood, IL: Scientific Software International, Inc.

الملاحق

أ

مستوى الفقرة	طول الاختبار عندما يكون حجم العينة 100 فرد								
	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-4.000	0.047	0.163	0.290	0.034	0.461	0.074	0.014	0.415	0.034
-3.875	0.065	0.178	0.364	0.043	0.502	0.086	0.037	0.459	0.081
-3.750	0.080	0.195	0.413	0.049	0.547	0.089	0.048	0.508	0.094
-3.625	0.116	0.214	0.543	0.057	0.595	0.096	0.066	0.560	0.118
-3.500	0.162	0.237	0.685	0.072	0.648	0.112	0.090	0.618	0.145
-3.375	0.221	0.262	0.844	0.094	0.704	0.133	0.120	0.680	0.176
-3.250	0.315	0.291	1.084	0.125	0.765	0.164	0.165	0.747	0.221
-3.125	0.487	0.323	1.508	0.178	0.830	0.215	0.226	0.819	0.277
-3.000	0.779	0.360	2.166	0.263	0.899	0.292	0.314	0.896	0.350
-2.875	1.322	0.401	3.300	0.401	0.972	0.412	0.423	0.978	0.433
-2.750	2.290	0.446	5.136	0.641	1.049	0.611	0.560	1.066	0.526
-2.625	3.682	0.496	7.431	0.975	1.130	0.863	0.718	1.159	0.619
-2.500	5.186	0.550	9.435	1.401	1.214	1.154	0.904	1.258	0.719
-2.375	7.007	0.608	11.527	1.856	1.300	1.427	1.140	1.362	0.837
-2.250	8.817	0.670	13.159	2.210	1.389	1.591	1.432	1.472	0.973
-2.125	10.032	0.735	13.642	2.427	1.479	1.641	1.768	1.587	1.114
-2.000	10.792	0.803	13.433	2.533	1.570	1.614	2.092	1.707	1.225
-1.875	11.403	0.873	13.061	2.601	1.660	1.567	2.414	1.832	1.318
-1.750	11.887	0.943	12.601	2.643	1.749	1.511	2.698	1.961	1.376
-1.625	12.244	1.013	12.083	2.663	1.836	1.450	2.878	2.093	1.375
-1.500	12.394	1.082	11.454	2.682	1.920	1.397	3.019	2.228	1.355
-1.375	12.463	1.148	10.854	2.706	2.001	1.353	3.129	2.365	1.323
-1.250	12.525	1.211	10.339	2.735	2.077	1.317	3.215	2.501	1.286
-1.125	12.600	1.271	9.918	2.771	2.148	1.290	3.282	2.636	1.245
-1.000	12.720	1.325	9.599	2.812	2.213	1.271	3.333	2.767	1.205

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 100 فرد: 20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-0.875	12.843	1.375	9.343	2.855	2.271	1.257	3.387	2.892	1.171
-0.750	13.188	1.419	9.295	2.898	2.323	1.248	3.439	3.010	1.142
-0.625	13.815	1.457	9.480	2.992	2.367	1.264	3.490	3.119	1.119
-0.500	14.443	1.490	9.692	3.129	2.404	1.302	3.536	3.217	1.099
-0.375	14.619	1.517	9.636	3.254	2.434	1.337	3.579	3.302	1.084
-0.250	14.435	1.538	9.383	3.326	2.455	1.355	3.623	3.373	1.074
-0.125	13.996	1.554	9.008	3.368	2.469	1.364	3.642	3.430	1.062
0.000	13.267	1.563	8.488	3.382	2.475	1.366	3.615	3.471	1.041
0.125	12.441	1.567	7.942	3.374	2.474	1.364	3.533	3.497	1.010
0.250	11.766	1.564	7.522	3.346	2.465	1.357	3.415	3.508	0.973
0.375	11.164	1.556	7.175	3.305	2.449	1.350	3.269	3.504	0.933
0.500	10.531	1.542	6.828	3.245	2.427	1.338	3.075	3.486	0.882
0.625	9.991	1.523	6.559	3.134	2.398	1.307	2.839	3.454	0.822
0.750	9.297	1.499	6.202	2.913	2.362	1.233	2.554	3.410	0.749
0.875	8.624	1.470	5.866	2.607	2.322	1.123	2.240	3.355	0.668
1.000	8.118	1.437	5.650	2.242	2.276	0.985	1.918	3.289	0.583
1.125	7.421	1.400	5.302	1.885	2.226	0.847	1.651	3.215	0.513
1.250	6.753	1.359	4.968	1.566	2.172	0.721	1.426	3.132	0.455
1.375	6.223	1.316	4.729	1.311	2.114	0.620	1.232	3.044	0.405
1.500	5.784	1.270	4.555	1.114	2.054	0.543	1.079	2.950	0.366
1.625	5.541	1.222	4.533	0.948	1.991	0.476	0.958	2.851	0.336
1.750	5.263	1.173	4.486	0.796	1.926	0.413	0.870	2.749	0.316
1.875	5.008	1.123	4.458	0.663	1.860	0.356	0.794	2.646	0.300
2.000	4.787	1.073	4.462	0.558	1.794	0.311	0.709	2.540	0.279
2.125	4.592	1.022	4.493	0.459	1.727	0.266	0.620	2.435	0.255
2.250	4.433	0.972	4.561	0.366	1.660	0.220	0.532	2.330	0.228

مستوى القدرة	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
طول الاختبار عندما يكون حجم العينة 100 فرد									
2.375	4.302	0.922	4.664	0.266	1.593	0.167	0.449	2.225	0.202
2.500	4.215	0.874	4.824	0.186	1.528	0.122	0.358	2.122	0.169
2.625	4.077	0.826	4.937	0.125	1.463	0.085	0.274	2.021	0.136
2.750	3.905	0.780	5.009	0.084	1.400	0.060	0.202	1.923	0.105
2.875	3.765	0.735	5.126	0.058	1.338	0.044	0.146	1.827	0.080
3.000	3.588	0.691	5.191	0.042	1.278	0.033	0.102	1.734	0.059
3.125	3.221	0.649	4.961	0.031	1.220	0.026	0.071	1.645	0.043
3.250	2.613	0.609	4.290	0.025	1.163	0.021	0.052	1.558	0.034
3.375	1.841	0.571	3.224	0.020	1.109	0.018	0.038	1.476	0.026
3.500	1.087	0.535	2.033	0.017	1.056	0.016	0.028	1.396	0.020
3.625	0.674	0.500	1.349	0.014	1.005	0.014	0.021	1.320	0.016
3.750	0.374	0.467	0.802	0.012	0.957	0.012	0.015	1.248	0.012
3.875	0.279	0.436	0.640	0.011	0.910	0.012	0.010	1.179	0.008

ب

مستوى القدرة	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
طول الاختبار عندما يكون حجم العينة 250 فرداً									
-4.000	0.150	0.076	1.984	0.006	0.230	0.025	0.067	0.218	0.308
-3.875	0.180	0.087	2.082	0.008	0.252	0.034	0.075	0.250	0.299
-3.750	0.231	0.100	2.320	0.013	0.276	0.046	0.081	0.286	0.282
-3.625	0.306	0.115	2.675	0.022	0.303	0.073	0.093	0.328	0.283
-3.500	0.419	0.132	3.182	0.039	0.332	0.117	0.108	0.375	0.288
-3.375	0.599	0.151	3.956	0.068	0.365	0.187	0.124	0.429	0.290
-3.250	0.841	0.174	4.839	0.110	0.401	0.275	0.143	0.489	0.292
-3.125	1.156	0.199	5.799	0.175	0.440	0.398	0.164	0.557	0.294
-3.000	1.460	0.228	6.392	0.253	0.484	0.523	0.189	0.634	0.298

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 250 فرداً: 20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-2.875	1.813	0.261	6.943	0.336	0.533	0.630	0.222	0.719	0.308
-2.750	2.037	0.298	6.848	0.419	0.587	0.714	0.264	0.814	0.324
-2.625	2.245	0.338	6.643	0.513	0.646	0.794	0.317	0.919	0.345
-2.500	2.482	0.383	6.488	0.620	0.711	0.873	0.380	1.034	0.368
-2.375	2.747	0.432	6.366	0.746	0.782	0.954	0.450	1.159	0.388
-2.250	3.035	0.485	6.261	0.885	0.859	1.030	0.525	1.296	0.405
-2.125	3.349	0.542	6.175	1.020	0.943	1.081	0.595	1.443	0.412
-2.000	3.643	0.604	6.029	1.214	1.034	1.175	0.656	1.602	0.409
-1.875	3.915	0.670	5.840	1.417	1.130	1.254	0.720	1.772	0.406
-1.750	4.135	0.741	5.581	1.631	1.233	1.323	0.778	1.953	0.399
-1.625	4.323	0.816	5.300	1.831	1.341	1.365	0.828	2.144	0.386
-1.500	4.420	0.895	4.939	2.024	1.453	1.393	0.886	2.347	0.378
-1.375	4.540	0.978	4.640	2.201	1.570	1.402	0.941	2.560	0.367
-1.250	4.812	1.066	4.514	2.337	1.689	1.384	1.000	2.782	0.359
-1.125	5.121	1.158	4.424	2.437	1.810	1.346	1.057	3.014	0.351
-1.000	5.477	1.252	4.374	2.511	1.932	1.300	1.113	3.253	0.342
-0.875	5.808	1.349	4.304	2.578	2.054	1.255	1.174	3.500	0.336
-0.750	6.069	1.448	4.192	2.634	2.175	1.211	1.243	3.750	0.332
-0.625	6.371	1.546	4.121	2.685	2.293	1.171	1.314	4.004	0.328
-0.500	6.636	1.642	4.042	2.728	2.408	1.133	1.374	4.256	0.323
-0.375	6.871	1.732	3.967	2.767	2.518	1.099	1.413	4.505	0.314
-0.250	7.006	1.815	3.859	2.801	2.622	1.068	1.433	4.745	0.302
-0.125	7.009	1.888	3.712	2.829	2.719	1.041	1.431	4.972	0.288
0.000	6.886	1.948	3.535	2.839	2.807	1.012	1.412	5.181	0.273
0.125	6.681	1.993	3.353	2.819	2.884	0.977	1.380	5.366	0.257
0.250	6.378	2.021	3.156	2.754	2.950	0.933	1.329	5.523	0.241

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 250 فرداً: 20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
0.375	5.901	2.032	2.904	2.641	3.004	0.879	1.255	5.646	0.222
0.500	5.204	2.025	2.570	2.494	3.042	0.820	1.161	5.732	0.203
0.625	4.536	2.002	2.266	2.330	3.066	0.760	1.061	5.777	0.184
0.750	3.979	1.964	2.026	2.146	3.073	0.698	0.966	5.781	0.167
0.875	3.514	1.912	1.838	1.927	3.065	0.629	0.872	5.744	0.152
1.000	3.324	1.850	1.797	1.712	3.040	0.563	0.788	5.667	0.139
1.125	3.245	1.779	1.824	1.533	3.001	0.511	0.710	5.553	0.128
1.250	3.182	1.702	1.869	1.376	2.947	0.467	0.634	5.406	0.117
1.375	3.159	1.622	1.948	1.214	2.881	0.422	0.552	5.231	0.106
1.500	3.130	1.539	2.034	1.048	2.803	0.374	0.463	5.032	0.092
1.625	3.088	1.456	2.122	0.890	2.717	0.328	0.386	4.814	0.080
1.750	3.046	1.373	2.219	0.740	2.622	0.282	0.319	4.584	0.070
1.875	2.982	1.292	2.308	0.598	2.522	0.237	0.260	4.344	0.060
2.000	2.859	1.213	2.357	0.467	2.418	0.193	0.212	4.100	0.052
2.125	2.665	1.137	2.344	0.357	2.312	0.154	0.173	3.855	0.045
2.250	2.383	1.064	2.239	0.263	2.204	0.120	0.141	3.613	0.039
2.375	1.981	0.995	1.992	0.191	2.096	0.091	0.116	3.376	0.034
2.500	1.537	0.928	1.655	0.135	1.990	0.068	0.096	3.146	0.030
2.625	1.000	0.866	1.155	0.094	1.885	0.050	0.080	2.924	0.027
2.750	0.616	0.806	0.763	0.058	1.782	0.033	0.068	2.713	0.025
2.875	0.347	0.750	0.462	0.036	1.683	0.021	0.059	2.513	0.024
3.000	0.179	0.698	0.256	0.023	1.587	0.014	0.053	2.323	0.023
3.125	0.098	0.648	0.152	0.016	1.495	0.010	0.047	2.145	0.022
3.250	0.061	0.602	0.102	0.012	1.406	0.008	0.042	1.979	0.021
3.375	0.042	0.559	0.075	0.009	1.322	0.007	0.037	1.823	0.020
3.500	0.031	0.519	0.059	0.007	1.242	0.006	0.031	1.679	0.018

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 250 فرداً								
	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
3.625	0.025	0.481	0.051	0.006	1.165	0.005	0.025	1.545	0.016
3.750	0.021	0.446	0.046	0.005	1.093	0.005	0.018	1.421	0.012
3.875	0.018	0.413	0.043	0.005	1.025	0.005	0.014	1.306	0.011

ج

طول الاختبار عندما يكون حجم العينة 500 فرد

مستوى القدرة	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-4.000	0.001	0.103	0.010	0.003	0.098	0.034	0.008	0.353	0.290
-3.875	0.001	0.119	0.012	0.006	0.110	0.053	0.008	0.393	0.364
-3.750	0.002	0.137	0.014	0.009	0.123	0.070	0.009	0.439	0.413
-3.625	0.003	0.156	0.020	0.013	0.138	0.094	0.010	0.489	0.543
-3.500	0.005	0.176	0.027	0.019	0.154	0.124	0.012	0.544	0.685
-3.375	0.009	0.198	0.046	0.029	0.172	0.171	0.015	0.605	0.844
-3.250	0.021	0.221	0.095	0.048	0.193	0.250	0.021	0.673	1.084
-3.125	0.056	0.246	0.228	0.082	0.215	0.379	0.033	0.748	1.508
-3.000	0.163	0.272	0.599	0.135	0.241	0.563	0.055	0.830	2.166
-2.875	0.389	0.300	1.297	0.218	0.269	0.810	0.089	0.920	3.300
-2.750	0.634	0.330	1.924	0.342	0.302	1.134	0.129	1.020	5.136
-2.625	0.819	0.361	2.268	0.494	0.338	1.461	0.175	1.130	7.431
-2.500	0.954	0.395	2.416	0.681	0.379	1.796	0.219	1.251	9.435
-2.375	1.055	0.431	2.449	0.910	0.426	2.138	0.264	1.383	11.527
-2.250	1.157	0.470	2.462	1.116	0.478	2.335	0.316	1.528	13.159
-2.125	1.272	0.513	2.481	1.271	0.536	2.370	0.379	1.687	13.642
-2.000	1.398	0.559	2.501	1.400	0.602	2.326	0.450	1.859	13.433
-1.875	1.553	0.610	2.545	1.536	0.675	2.275	0.513	2.047	13.061
-1.750	1.742	0.667	2.613	1.681	0.756	2.222	0.576	2.248	12.601
-1.625	1.959	0.729	2.688	1.798	0.846	2.126	0.638	2.465	12.083
-1.500	2.276	0.797	2.854	1.883	0.944	1.995	0.703	2.695	11.454
-1.375	2.711	0.873	3.104	1.940	1.050	1.847	0.764	2.937	10.854
-1.250	3.280	0.957	3.428	1.985	1.165	1.704	0.826	3.190	10.339
-1.125	3.921	1.049	3.739	2.026	1.288	1.573	0.893	3.451	9.918
-1.000	4.327	1.150	3.764	2.077	1.419	1.464	0.973	3.716	9.599
-0.875	4.554	1.259	3.616	2.162	1.557	1.389	1.057	3.982	9.343

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 500 فرد								
	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-0.750	4.656	1.378	3.380	2.288	1.702	1.345	1.135	4.245	9.295
-0.625	4.700	1.504	3.125	2.434	1.853	1.314	1.194	4.501	9.480
-0.500	4.725	1.636	2.888	2.564	2.009	1.276	1.230	4.746	9.692
-0.375	4.740	1.772	2.675	2.650	2.170	1.221	1.245	4.974	9.636
-0.250	4.741	1.909	2.484	2.682	2.336	1.148	1.246	5.182	9.383
-0.125	4.693	2.042	2.298	2.677	2.505	1.069	1.234	5.365	9.008
0.000	4.570	2.169	2.108	2.646	2.679	0.988	1.210	5.519	8.488
0.125	4.417	2.283	1.935	2.585	2.856	0.905	1.172	5.641	7.942
0.250	4.293	2.381	1.803	2.476	3.038	0.815	1.117	5.728	7.522
0.375	4.178	2.458	1.700	2.318	3.224	0.719	1.042	5.777	7.175
0.500	4.069	2.511	1.620	2.132	3.416	0.624	0.950	5.788	6.828
0.625	3.921	2.538	1.545	1.916	3.609	0.531	0.850	5.760	6.559
0.750	3.707	2.540	1.460	1.728	3.800	0.455	0.750	5.695	6.202
0.875	3.308	2.515	1.315	1.584	3.979	0.398	0.656	5.596	5.866
1.000	2.838	2.466	1.151	1.463	4.134	0.354	0.585	5.465	5.650
1.125	2.359	2.397	0.984	1.346	4.252	0.317	0.533	5.306	5.302
1.250	1.926	2.310	0.834	1.235	4.320	0.286	0.491	5.125	4.968
1.375	1.590	2.208	0.720	1.134	4.329	0.262	0.454	4.926	4.729
1.500	1.363	2.097	0.650	1.054	4.276	0.246	0.423	4.713	4.555
1.625	1.256	1.979	0.635	0.992	4.167	0.238	0.388	4.491	4.533
1.750	1.203	1.857	0.648	0.946	4.011	0.236	0.345	4.263	4.486
1.875	1.176	1.734	0.678	0.894	3.822	0.234	0.299	4.034	4.458
2.000	1.161	1.612	0.720	0.804	3.612	0.223	0.252	3.806	4.462
2.125	1.149	1.493	0.769	0.702	3.392	0.207	0.211	3.582	4.493
2.250	1.130	1.378	0.820	0.599	3.170	0.189	0.176	3.362	4.561
2.375	1.105	1.268	0.872	0.504	2.952	0.171	0.146	3.150	4.664

مستوى القدرة	طول الاختبار عندما يكون حجم العينة 500 فرد								
	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
2.500	1.063	1.164	0.914	0.424	2.741	0.155	0.120	2.945	4.824
2.625	0.956	1.065	0.898	0.349	2.540	0.138	0.098	2.749	4.937
2.750	0.820	0.973	0.843	0.248	2.349	0.106	0.080	2.562	5.009
2.875	0.693	0.887	0.781	0.159	2.168	0.073	0.067	2.385	5.126
3.000	0.562	0.807	0.696	0.095	1.999	0.048	0.057	2.217	5.191
3.125	0.460	0.734	0.626	0.057	1.840	0.031	0.049	2.059	4.961
3.250	0.392	0.666	0.589	0.037	1.692	0.022	0.044	1.910	4.290
3.375	0.351	0.604	0.582	0.027	1.554	0.017	0.040	1.770	3.224
3.500	0.320	0.547	0.585	0.021	1.425	0.015	0.037	1.639	2.033
3.625	0.307	0.495	0.622	0.018	1.306	0.014	0.035	1.517	1.349
3.750	0.279	0.447	0.624	0.014	1.196	0.012	0.033	1.403	0.802
3.875	0.241	0.404	0.596	0.012	1.094	0.011	0.030	1.297	0.640

د

طول الاختبار عندما يكون حجم العينة 1000 فرد

مستوى القدرة	20 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	20 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	20 فقرة: الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	40 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	40 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	40 فقرة: الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	60 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	60 فقرة: دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	60 فقرة: الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-4.000	0.464	0.033	14.246	0.005	0.069	0.075	0.010	0.085	0.113
-3.875	0.464	0.042	11.084	0.006	0.079	0.076	0.014	0.100	0.139
-3.750	0.472	0.054	8.825	0.008	0.090	0.094	0.019	0.117	0.160
-3.625	0.495	0.068	7.285	0.012	0.102	0.122	0.027	0.138	0.195
-3.500	0.531	0.086	6.194	0.022	0.117	0.190	0.043	0.163	0.263
-3.375	0.593	0.107	5.530	0.048	0.133	0.362	0.071	0.192	0.373
-3.250	0.699	0.133	5.266	0.119	0.152	0.781	0.114	0.226	0.505
-3.125	0.869	0.162	5.352	0.296	0.174	1.704	0.165	0.265	0.624
-3.000	1.117	0.196	5.691	0.541	0.198	2.733	0.218	0.312	0.700
-2.875	1.444	0.234	6.175	0.655	0.226	2.898	0.256	0.367	0.697
-2.750	1.713	0.275	6.230	0.693	0.258	2.689	0.287	0.430	0.666
-2.625	1.870	0.319	5.863	0.711	0.295	2.412	0.312	0.504	0.619
-2.500	1.913	0.365	5.240	0.734	0.337	2.179	0.333	0.590	0.564
-2.375	1.913	0.412	4.639	0.770	0.386	1.995	0.357	0.689	0.518
-2.250	1.919	0.460	4.167	0.827	0.442	1.869	0.389	0.803	0.484
-2.125	1.935	0.509	3.802	0.915	0.508	1.800	0.429	0.934	0.459
-2.000	1.960	0.558	3.513	1.019	0.585	1.742	0.472	1.082	0.436
-1.875	1.987	0.608	3.268	1.111	0.675	1.648	0.508	1.250	0.406
-1.750	2.013	0.660	3.050	1.179	0.778	1.515	0.537	1.439	0.373
-1.625	2.035	0.715	2.846	1.213	0.898	1.352	0.566	1.649	0.343
-1.500	2.112	0.775	2.725	1.233	1.034	1.193	0.597	1.882	0.317
-1.375	2.277	0.841	2.706	1.254	1.187	1.056	0.634	2.137	0.297
-1.250	2.416	0.916	2.639	1.286	1.356	0.948	0.676	2.416	0.280
-1.125	2.539	1.000	2.539	1.358	1.540	0.882	0.720	2.717	0.265
-1.000	2.665	1.096	2.431	1.474	1.736	0.850	0.770	3.041	0.253

مستوى القدرة	20 فقرة			40 فقرة			60 فقرة		
	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة اللابارامترية	دالة معلومات الاختبار وفقاً لنظرية استجابة الفقرة البارامترية	الكفاءة النسبية لدالة معلومات اللابارامترية إلى البارامترية
-0.875	2.793	1.205	2.318	1.595	1.939	0.823	0.824	3.388	0.243
-0.750	2.945	1.327	2.219	1.707	2.146	0.796	0.885	3.757	0.235
-0.625	3.097	1.463	2.117	1.799	2.353	0.765	0.938	4.148	0.226
-0.500	3.234	1.612	2.006	1.863	2.556	0.729	0.978	4.561	0.215
-0.375	3.337	1.772	1.883	1.900	2.752	0.690	0.998	4.993	0.200
-0.250	3.426	1.940	1.766	1.917	2.940	0.652	1.002	5.442	0.184
-0.125	3.521	2.112	1.667	1.918	3.117	0.615	0.994	5.899	0.168
0.000	3.596	2.283	1.575	1.904	3.282	0.580	0.976	6.357	0.153
0.125	3.543	2.447	1.448	1.872	3.435	0.545	0.945	6.800	0.139
0.250	3.342	2.597	1.287	1.809	3.575	0.506	0.901	7.212	0.125
0.375	3.056	2.726	1.121	1.702	3.701	0.460	0.839	7.574	0.111
0.500	2.771	2.829	0.979	1.567	3.811	0.411	0.764	7.870	0.097
0.625	2.524	2.901	0.870	1.424	3.903	0.365	0.680	8.084	0.084
0.750	2.290	2.939	0.779	1.292	3.975	0.325	0.602	8.208	0.073
0.875	2.063	2.943	0.701	1.165	4.025	0.290	0.528	8.241	0.064
1.000	1.849	2.912	0.635	1.051	4.051	0.260	0.453	8.190	0.055
1.125	1.663	2.851	0.583	0.940	4.052	0.232	0.381	8.068	0.047
1.250	1.478	2.762	0.535	0.823	4.027	0.204	0.319	7.889	0.040
1.375	1.282	2.649	0.484	0.706	3.977	0.178	0.269	7.668	0.035
1.500	1.032	2.518	0.410	0.588	3.902	0.151	0.229	7.413	0.031
1.625	0.777	2.373	0.327	0.474	3.805	0.125	0.196	7.124	0.028
1.750	0.595	2.219	0.268	0.382	3.687	0.104	0.169	6.793	0.025
1.875	0.489	2.060	0.238	0.303	3.552	0.085	0.146	6.414	0.023
2.000	0.419	1.899	0.220	0.240	3.403	0.071	0.127	5.986	0.021
2.125	0.369	1.741	0.212	0.185	3.243	0.057	0.109	5.517	0.020
2.250	0.332	1.587	0.209	0.141	3.075	0.046	0.090	5.026	0.018

طول الاختبار عندما يكون حجم العينة 1000 فرد

Abstract

Al-qise. Hussen Abdelnabi. Accuracy of Item Parameters and Ability Estimated Parametric and Non-parametric Models of Item Response Theory

. PhD Dissertation. Yarmouk University. 2013. (Supervisor: Prof. Sari Saleem Sawaqed).

The purpose of the study is to compar the accuracy of Item and ability Parameters by using parametric and Non-parametric item response theory models within the difference in sample's size and test length Depending on accuracy and bias standards as well as root Mean square error (RMSE) standard. To achieve the aims of this study the researcher generated the abilities of samples sized 100, 250, 500, 1000 from a normal distribution with a mean of (0) and standard deviation of (1). Depending on ability feature tests were generated according to the following classification (20,40,60) items under the normal distribution of item difficulty with a mean of (0) and standards deviation of (1) as well as regular distribution for discrimination feature with a primary value of (0.4) and final value of (1.2) as well as guessing feature regular distribution with a primary value of (0.2) and final value of (0.3) proposing multiple choice tests with four alternatives by using WinGen program for data generating.

To answer the questions of the study the researcher used BILOG-MG software to evaluate the items features and ability through parametric Marginal Maximum likelihood estimation (MML) then comparing the generated features with the real ones by using Bias and RMSE indicators. Pearson Correlation coefficient was used as a trust indicator between the real and generated features. TESTGRAF software was used to evaluate the item features and ability through softening the core (KS) then all were

evaluated by using Bias and RMSE indicators. Pearson coefficient was used as a trust indicator between the real and generated features.

The findings of the study showed that the values of correlations between the real ability feature(θ) and the ability feature(θ) estimated through the parametric and non-parametric methods was positive and significant statically and can be classified as high. Moreover, the findings showed that the correlation coefficient values between the real features (a,b,c) and the features evaluated through the parametric method were positive and significant statically and can be classified as high and it is higher than those evaluated through the non-parametric method in all the cases in this study. With regard to the accuracy indicators in the measurement, the findings showed that the bias and RMSE indicators of the parametric method were lower than those in the non-parametric method in all cases. Furthermore, 3-way variance analysis for repeated measures for items feature (a) and ability feature (θ) showed that there are significant statistical differences at the level of (α =0.05) between the means of bias indicator values for items feature (a) and ability feature(θ) attributed to the interaction of the used model (parametric, non-parametric) with sample size and test length variables. -way variance analysis for repeated measures for the evaluations of (c, b) items showed that there are no significant statistical differences at the level of (α =0.05) between the means of bias indicator values for items feature (a) and ability feature(θ) attributed to the interaction of the used model (parametric, non-parametric) with sample size and test length variables.

With regard to RMSE the findings showed that the means of values for the parametric method regarding (a, b, c) and ability feature (θ) were less than those in the non-parametric method in all cases and this difference decreases with the increase of test length and sample size.

With regard to the compatibility as a trust indicator, the findings showed that the compatibility between the means of ability feature(θ) for the test items with generated data according to the parametric method and the real ability features were higher than the compatibility generated according to the non-parametric method.

With regard to the extent of compatibility as a trust indicator for the items features the findings showed that the comparability between the mean of item features of (a, b, c) with generated data according to the parametric method and the values of item features evaluations were higher than those generated according to the non-parametric method.

Key words: Evaluation accuracy. Item Features. Parametric and non-parametric Response Theory Models.